(51) **International Patent Classification[7]:** C12Q

(21) **International Application Number:** PCT/US02/14341

(22) **International Filing Date:** 2 May 2002 (02.05.2002)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**

60/288,935	3 May 2001 (03.05.2001)	US
60/289,087	7 May 2001 (07.05.2001)	US
60/289,620	8 May 2001 (08.05.2001)	US
60/289,621	8 May 2001 (08.05.2001)	US
60/289,817	9 May 2001 (09.05.2001)	US
60/289,818	9 May 2001 (09.05.2001)	US
60/290,194	11 May 2001 (11.05.2001)	US
60/290,753	14 May 2001 (14.05.2001)	US
60/291,189	15 May 2001 (15.05.2001)	US
60/291,243	16 May 2001 (16.05.2001)	US
60/292,001	18 May 2001 (18.05.2001)	US
60/292,374	21 May 2001 (21.05.2001)	US
60/292,587	22 May 2001 (22.05.2001)	US
60/293,107	23 May 2001 (23.05.2001)	US
60/293,589	24 May 2001 (24.05.2001)	US
60/293,747	25 May 2001 (25.05.2001)	US
60/294,110	29 May 2001 (29.05.2001)	US
60/294,434	30 May 2001 (30.05.2001)	US
60/312,192	14 August 2001 (14.08.2001)	US
60/313,173	17 August 2001 (17.08.2001)	US
60/313,187	17 August 2001 (17.08.2001)	US
60/318,728	12 September 2001 (12.09.2001)	US
60/318,744	12 September 2001 (12.09.2001)	US
60/335,910	15 November 2001 (15.11.2001)	US
60/333,942	28 November 2001 (28.11.2001)	US
60/333,891	28 November 2001 (28.11.2001)	US
60/345,776	3 January 2002 (03.01.2002)	US
60/345,220	4 January 2002 (04.01.2002)	US
10/136,071	1 May 2002 (01.05.2002)	US

(63) **Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:**

US	60/288,935 (CON)
Filed on	3 May 2001 (03.05.2001)
US	60/289,087 (CON)
Filed on	7 May 2001 (07.05.2001)

US	60/318,744 (CON)
Filed on	12 September 2001 (12.09.2001)
US	60/313,187 (CON)
Filed on	17 August 2001 (17.08.2001)
US	60/333,891 (CON)
Filed on	28 November 2001 (28.11.2001)
US	60/345,776 (CON)
Filed on	3 January 2002 (03.01.2002)
US	60/312,192 (CON)
Filed on	14 August 2001 (14.08.2001)
US	60/289,621 (CON)
Filed on	8 May 2001 (08.05.2001)
US	60/289,818 (CON)
Filed on	9 May 2001 (09.05.2001)
US	60/289,817 (CON)
Filed on	9 May 2001 (09.05.2001)
US	60/318,728 (CON)
Filed on	12 September 2001 (12.09.2001)
US	60/289,620 (CON)
Filed on	8 May 2001 (08.05.2001)
US	60/290,194 (CON)
Filed on	11 May 2001 (11.05.2001)
US	60/333,942 (CON)
Filed on	28 November 2001 (28.11.2001)
US	60/290,753 (CON)
Filed on	14 May 2001 (14.05.2001)
US	60/291,189 (CON)
Filed on	15 May 2001 (15.05.2001)
US	60/335,910 (CON)
Filed on	15 November 2001 (15.11.2001)
US	60/313,173 (CON)
Filed on	17 August 2001 (17.08.2001)
US	60/291,243 (CON)
Filed on	16 May 2001 (16.05.2001)
US	60/292,001 (CON)
Filed on	18 May 2001 (18.05.2001)
US	60/292,374 (CON)
Filed on	21 May 2001 (21.05.2001)
US	60/345,220 (CON)
Filed on	4 January 2002 (04.01.2002)
US	60/292,587 (CON)
Filed on	22 May 2001 (22.05.2001)
US	60/293,107 (CON)
Filed on	23 May 2001 (23.05.2001)
US	60/293,589 (CON)
Filed on	24 May 2001 (24.05.2001)

(54) **Title:** THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS OF USE

(57) **Abstract:** Disclosed herein are nucleic acid sequences that encode G-coupled protein-receptor related polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies, which immunospecifically-bind to the-polypeptide, as well as derivatives, variants, mutants, or fragments of the aforementioned polypeptide, polynucleotide, or antibody. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.

US	60/294,110 (CON)
Filed on	29 May 2001 (29.05.2001)
US	60/293,747 (CON)
Filed on	25 May 2001 (25.05.2001)
US	60/294,434 (CON)
Filed on	30 May 2001 (30.05.2001)
US	Not furnished (CON)
Filed on	1 May 2002 (01.05.2002)

(71) **Applicant** *(for all designated States except US)*: **CURA-GEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th floor, New Haven, CT 06511 (US).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **ALSOBROOK, John, P., II** [US/US]; 60 Lake Drive, Madison, CT 06443 (US). **ANDERSON, David, W.** [US/US]; 85 Montoya Drive, Branford, CT 06405 (US). **BOLDOG, Ferenc, L.** [HU/US]; 1687 Hartford Turnpike, North Haven, CT 06473 (US). **BURGESS, Catherine, E.** [US/US]; 90 Carriage Hill Drive, Weathersfield, CT 06109 (US). **CASMAN, Stacie, J.** [US/US]; 17 Peck Street, North Haven, CT 06473 (US). **EDINGER, Schlomit, R.** [US/US]; 766 Edgewood Avenue, New Haven, CT 06515 (US). **ELLERMAN, Karen** [US/US]; 87 Montoya Drive, Branford, CT 06405 (US). **GANGOLLI, Esha, A.** [IN/US]; 31 Strawberry Hill Road, Madison, CT 06443 (US). **GERLACH, Valerie, L.** [US/US]; 18 Rock Pasture Road, Branford, CT 06405 (US). **GORMAN, Linda** [US/US]; 329 Monticello Drive, Branford, CT 06405 (US). **GUNTHER, Erik** [US/US]; 34 Bryan Road, Branford, CT 06405 (US). **HERRMANN, John, L.** [US/US]; 78 Barnshed Lane, Guilford, CT 06437 (US). **JI, Weizhen** [CN/US]; 3 Business Park Drive, Room 101, Branford, CT 06405 (US). **LEPLEY, Denise, M.** [US/US]; 51 Church Street, Branford, CT 06405 (US). **LEWIN, David, A.** [US/US]; 298 Lawrence Street, New Haven, CT 06511 (US). **LI, Li** [CN/US]; 56 Jerimoth Lane, Branford, CT 06405 (US). **MACDOUGALL, John, R.** [CA/US]; 117 Russell Street, Hamden, CT 06517 (US). **MALYANKAR, Uriel, M.** [IN/US]; 229 Branford Road, number 330, Branford, CT 06405 (US). **MEZES, Peter, D.** [CA/US]; 7 Clark's Lane, Old Lyme, CT 06371 (US). **PADIGARU, Muralidhara** [IN/US]; 71 Hampton Park, Branford, CT 06405 (US). **PATTURAJAN, Meera** [IN/US]; 45 Harrison Avenue, Apartment 1C, Branford, CT 06405 (US). **PEYMAN, John, A.** [US/US]; 336 West Rock Avenue, New Haven, CT 06515 (US). **RASTELLI, Luca** [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437 (US). **RIEGER, Daniel,**

K. [DE/US]; 10A McKinnel Court, Branford, CT 06405 (US). **ROTHENBERG, Mark, E.** [US/US]; 2 Allen Road, Clinton, CT 06413 (US). **SHENOY, Suresh, G.** [IN/US]; 15 Millwood Drive, Branford, CT 06405 (US). **SMITHSON, Glennda** [US/US]; 125 Michael Drive, Guilford, CT 06435 (US). **SPYTEK, Kimberly, A.** [US/US]; 28 Court Street, number 1, New Haven, CT 06511 (US). **STONE, David, J.** [US/US]; 223 Whitehorn Drive, Guilford, CT 06437 (US). **TAUPIER, Raymond, J., Jr.** [US/US]; 34 Pardee Place Extension, East Haven, CT 06512 (US). **TCHERNEV, Velizar, T.** [BG/US]; 45 Jefferson Road, 3-12, Branford, CT 06405 (US). **VERNET, Corine, A., M.** [FR/US]; 1739 Foxon Road, Apt. L6, Branford, CT 06471 (US). **VOSS, Edward, Z.** [US/US]; 123 Knollwood Drive, Wallingford, CT 06492 (US). **ZERHUSEN, Bryan, D.** [US/US]; 337 Monticello Drive, Branford, CT 06405 (US). **ZHONG, Haihong** [CN/US]; 2269 Long Hill Road, Guilford, CT 06437 (US). **MILLER, Charles, E.** [US/US]; 98 Saddle Hill Drive, Guilford, CT 06437 (US).

(74) **Agent: ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky & Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS OF USE

FIELD OF THE INVENTION

5   The present invention relates to novel polypeptides, and the nucleic acids encoding them, having properties related to stimulation of biochemical or physiological responses in a cell, a tissue, an organ or an organism. More particularly, the novel polypeptides are gene products of novel genes, or are specified biologically active fragments or derivatives thereof. Methods of use encompass diagnostic and prognostic assay procedures as well as

10   methods of treating diverse pathological conditions.

BACKGROUND OF THE INVENTION

Eukaryotic cells are characterized by biochemical and physiological processes, which under normal conditions are exquisitely balanced to achieve the preservation and

5   propagation of the cells. When such cells are components of multicellular organisms such as vertebrates or, more particularly, organisms such as mammals, the regulation of the biochemical and physiological processes involves intricate signaling pathways. Frequently, such signaling pathways include constituted of extracellular signaling proteins, cellular receptors that bind the signaling proteins and signal transducing components

10   located within the cells.

Signaling proteins may be classified as endocrine effectors, paracrine effectors or autocrine effectors. Endocrine effectors are signaling molecules secreted by a given organ into the circulatory system, which are then transported to a distant target organ or tissue. The target cells include the receptors for the endocrine effector, and when the endocrine

15   effector binds, a signaling cascade is induced. Paracrine effectors involve secreting cells and receptor cells in close proximity to each other, such as two different classes of cells in the same tissue or organ. One class of cells secretes the paracrine effector, which then reaches the second class of cells, for example by diffusion through the extracellular fluid. The second class of cells contains the receptors for the paracrine effector; binding of the

20   effector results in induction of the signaling cascade that elicits the corresponding biochemical or physiological effect. Autocrine effectors are highly analogous to paracrine effectors, except that the same cell type that secretes the autocrine effector also contains the receptor. Thus the autocrine effector binds to receptors on the same cell, or on identical neighboring cells. The binding process then elicits the characteristic biochemical

25   or physiological effect.

Signaling processes may elicit a variety of effects on cells and tissues including, by way of nonlimiting example, induction of cell or tissue proliferation, suppression of growth or proliferation, induction of differentiation or maturation of a cell or tissue, and suppression of differentiation or maturation of a cell or tissue.

30   Many pathological conditions involve dysregulation of expression of important effector proteins. In certain classes of pathologies the dysregulation is manifested as diminished or suppressed level of synthesis and secretion of protein effectors. In other classes of pathologies the dysregulation is manifested as increased or up-regulated level of

synthesis and secretion of protein effectors. In a clinical setting a subject may be suspected of suffering from a condition brought on by altered or mis-regulated levels of a protein effector of interest. Therefore there is a need to assay for the level of the protein effector of interest in a biological sample from such a subject, and to compare the level

5   with that characteristic of a nonpathological condition. There also is a need to provide the protein effector as a product of manufacture. Administration of the effector to a subject in need thereof is useful in treatment of the pathological condition. Accordingly, there is a need for a method of treatment of a pathological condition brought on by a diminished or suppressed levels of the protein effector of interest. In addition, there is a need for a

10   method of treatment of a pathological condition brought on by a increased or up-regulated levels of the protein effector of interest.

SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of isolated polypeptides

15   including amino acid sequences selected from mature forms of the amino acid sequences selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62. The invention also is based in part upon variants of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62, wherein any amino acid in the mature form is changed to a

20   different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed. In another embodiment, the invention includes the amino acid sequences selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62. In another embodiment, the invention also comprises variants of the amino acid sequence selected from the group consisting of SEQ

25   ID NO:2n, wherein n is an integer between 1 and 62 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed. The invention also involves fragments of any of the mature forms of the amino acid sequences selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62, or any other

30   amino acid sequence selected from this group. The invention also comprises fragments from these groups in which up to 15% of the residues are changed.

In another embodiment, the invention encompasses polypeptides that are naturally occurring allelic variants of the sequence selected from the group consisting of SEQ ID

NO:2n, wherein n is an integer between 1 and 62. These allelic variants include amino acid sequences that are the translations of nucleic acid sequences differing by a single nucleotide from nucleic acid sequences selected from the group consisting of SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 62. The variant polypeptide where any amino acid changed in the chosen sequence is changed to provide a conservative substitution.

In another embodiment, the invention comprises a pharmaceutical composition involving a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62 and a pharmaceutically acceptable carrier. In another embodiment, the invention involves a kit, including, in one or more containers, this pharmaceutical composition.

In another embodiment, the invention includes the use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease being selected from a pathology associated with a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62 wherein said therapeutic is the polypeptide selected from this group.

In another embodiment, the invention comprises a method for determining the presence or amount of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62 in a sample, the method involving providing the sample; introducing the sample to an antibody that binds immunospecifically to the polypeptide; and determining the presence or amount of antibody bound to the polypeptide, thereby determining the presence or amount of polypeptide in the sample.

In another embodiment, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62 in a first mammalian subject, the method involving measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and comparing the amount of the polypeptide in this sample to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, the disease, wherein an alteration in the

expression level of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

In another embodiment, the invention involves a method of identifying an agent that binds to a polypeptide with an amino acid sequence selected from the group

5   consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62, the method including introducing the polypeptide to the agent; and determining whether the agent binds to the polypeptide. The agent could be a cellular receptor or a downstream effector.

In another embodiment, the invention involves a method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to

10   aberrant expression or aberrant physiological interactions of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62, the method including providing a cell expressing the polypeptide of the invention and having a property or function ascribable to the polypeptide; contacting the cell with a composition comprising a candidate substance; and

15   determining whether the substance alters the property or function ascribable to the polypeptide; whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition devoid of the substance, the substance is identified as a potential therapeutic agent.

In another embodiment, the invention involves a method for screening for a

20   modulator of activity or of latency or predisposition to a pathology associated with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62, the method including administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of the invention, wherein the test animal recombinantly expresses the

25   polypeptide of the invention; measuring the activity of the polypeptide in the test animal after administering the test compound; and comparing the activity of the protein in the test animal with the activity of the polypeptide in a control animal not administered the polypeptide, wherein a change in the activity of the polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of, or

30   predisposition to, a pathology associated with the polypeptide of the invention. The recombinant test animal could express a test protein transgene or express the transgene under the control of a promoter at an increased level relative to a wild-type test animal The promoter may or may not b the native gene promoter of the transgene.

In another embodiment, the invention involves a method for modulating the
activity of a polypeptide with an amino acid sequence selected from the group consisting
of SEQ ID NO:2n, wherein n is an integer between 1 and 62, the method including
introducing a cell sample expressing the polypeptide with a compound that binds to the
5   polypeptide in an amount sufficient to modulate the activity of the polypeptide.

In another embodiment, the invention involves a method of treating or preventing a
pathology associated with a polypeptide with an amino acid sequence selected from the
group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62, the method
including administering the polypeptide to a subject in which such treatment or prevention
10   is desired in an amount sufficient to treat or prevent the pathology in the subject. The
subject could be human.

In another embodiment, the invention involves a method of treating a pathological
state in a mammal, the method including administering to the mammal a polypeptide in an
amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a
15   polypeptide having an amino acid sequence at least 95% identical to a polypeptide having
the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n
is an integer between 1 and 62 or a biologically active fragment thereof.

In another embodiment, the invention involves an isolated nucleic acid molecule
comprising a nucleic acid sequence encoding a polypeptide having an amino acid
20   sequence selected from the group consisting of a mature form of the amino acid sequence
given SEQ ID NO:2n, wherein n is an integer between 1 and 62; a variant of a mature
form of the amino acid sequence selected from the group consisting of SEQ ID NO:2n,
wherein n is an integer between 1 and 62 wherein any amino acid in the mature form of
the chosen sequence is changed to a different amino acid, provided that no more than 15%
25   of the amino acid residues in the sequence of the mature form are so changed; the amino
acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an
integer between 1 and 62; a variant of the amino acid sequence selected from the group
consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62, in which any
amino acid specified in the chosen sequence is changed to a different amino acid, provided
30   that no more than 15% of the amino acid residues in the sequence are so changed; a
nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino
acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an
integer between 1 and 62 or any variant of the polypeptide wherein any amino acid of the

chosen sequence is changed to a different amino acid, provided that no more than 10% of
the amino acid residues in the sequence are so changed; and the complement of any of the
nucleic acid molecules.

In another embodiment, the invention comprises an isolated nucleic acid molecule
having a nucleic acid sequence encoding a polypeptide comprising an amino acid
sequence selected from the group consisting of a mature form of the amino acid sequence
given SEQ ID NO:2n, wherein n is an integer between 1 and 62, wherein the nucleic acid
molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid
variant.

In another embodiment, the invention involves an isolated nucleic acid molecule
including a nucleic acid sequence encoding a polypeptide having an amino acid sequence
selected from the group consisting of a mature form of the amino acid sequence given
SEQ ID NO:2n, wherein n is an integer between 1 and 62 that encodes a variant
polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally
occurring polypeptide variant.

In another embodiment, the invention comprises an isolated nucleic acid molecule
having a nucleic acid sequence encoding a polypeptide comprising an amino acid
sequence selected from the group consisting of a mature form of the amino acid sequence
given SEQ ID NO:2n, wherein n is an integer between 1 and 62, wherein the nucleic acid
molecule differs by a single nucleotide from a nucleic acid sequence selected from the
group consisting of SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 62.

In another embodiment, the invention includes an isolated nucleic acid molecule
having a nucleic acid sequence encoding a polypeptide including an amino acid sequence
selected from the group consisting of a mature form of the amino acid sequence given
SEQ ID NO:2n, wherein n is an integer between 1 and 62, wherein the nucleic acid
molecule comprises a nucleotide sequence selected from the group consisting of the
nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is
an integer between 1 and 62; a nucleotide sequence wherein one or more nucleotides in the
nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is
an integer between 1 and 62 is changed from that selected from the group consisting of the
chosen sequence to a different nucleotide provided that no more than 15% of the
nucleotides are so changed; a nucleic acid fragment of the sequence selected from the
group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 62; and a

7

nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 62 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

5   In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 62, wherein the nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from

10   the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 62, or a complement of the nucleotide sequence.

           In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given

15   SEQ ID NO:2n, wherein n is an integer between 1 and 62, wherein the nucleic acid molecule has a nucleotide sequence in which any nucleotide specified in the coding sequence of the chosen nucleotide sequence is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides in the chosen coding sequence are so changed, an isolated second

20   polynucleotide that is a complement of the first polynucleotide, or a fragment of any of them.

           In another embodiment, the invention includes a vector involving the nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence

25   given SEQ ID NO:2n, wherein n is an integer between 1 and 62. This vector can have a promoter operably linked to the nucleic acid molecule. This vector can be located within a cell.

           In another embodiment, the invention involves a method for determining the presence or amount of a nucleic acid molecule having a nucleic acid sequence encoding a

30   polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 62 in a sample, the method including providing the sample; introducing the sample to a probe that binds to the nucleic acid molecule; and determining the presence or

8

amount of the probe bound to the nucleic acid molecule, thereby determining the presence or amount of the nucleic acid molecule in the sample. The presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type. The cell type can be cancerous.

5   In another embodiment, the invention involves a method for determining the presence of or predisposition for a disease associated with altered levels of a nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 62 in a first mammalian

10   subject, the method including measuring the amount of the nucleic acid in a sample from the first mammalian subject; and comparing the amount of the nucleic acid in the sample of step (a) to the amount of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease; wherein an alteration in the level of the nucleic acid in the first subject as compared to the control

15   sample indicates the presence of or predisposition to the disease.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable

20   methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

25   Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded

30   thereby. Included in the invention are the novel nucleic acid sequences, their encoded polypeptides, antibodies, and other related compounds. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX

9

proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel

sequences disclosed herein. Table 1 provides a summary of the NOVX nucleic acids and

their encoded polypeptides.

TABLE 1. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (amino acid)	Homology
NOV1a	CG100041-01	1	2	Trypsin like homo sapiens
NOV2a	CG105716-01	3	4	Germline oligomeric matrix protein
NOV3a	CG113569-01	5	6	Neuromedin U25 like homo sapiens
NOV3b	CG113569-03	7	8	Neuromedin U25 like homo sapiens
NOV4a	CG56602-02	9	10	Caldecrin like homo sapiens
NOV5a	CG57415-01	11	12	Neural cell adhesion protein like homo sapiens
NOV6a	CG58504-01	13	14	ADAMTS 12
NOV6b	169648407	15	16	ADAMTS 12
NOV6c	169648441	17	18	ADAMTS 12
NOV7a	CG58586-01	19	20	CASPR4 like homo sapiens
NOV7b	CG58586-02	21	22	CASPR4B like homo sapiens
NOV8a	CG93453-01	23	24	ADAMS-TS 3 precursor (KIAA0366) like homo sapiens
NOV8b	CG93453-02	25	26	ADAMS-TS 3 precursor
NOV8c	210387874	27	28	ADAMS-TS 3 precursor
NOV9a	CG95145-01	29	30	Gliacolin like homo sapiens
NOV10a	CG95250-01	31	32	Aminopeptidase N like homo sapiens
NOV10b	CG95250-02	33	34	Aminopeptidase N like homo sapiens
NOV10c	CG95250-03	35	36	Aminopeptidase N like homo sapiens
NOV11a	CG95430-01	37	38	Adiponectin like homo sapiens
NOV11b	CG95430-02	39	40	Adiponectin like homo sapiens
NOV11c	CG95430-03	41	42	Adiponectin like homo sapiens
NOV11d	CG95430-04	43	44	Adiponectin like homo sapiens
NOV11e	CG95430-06	45	46	Adiponectin like homo sapiens
NOV11f	175184045	47	48	Adiponectin like homo sapiens
NOV11g	175184049	49	50	Adiponectin like homo sapiens
NOV11h	175184053	51	52	Adiponectin like homo sapiens
NOV11i	175070796	53	54	Adiponectin like homo sapiens
NOV11j	175070804	55	56	Adiponectin like homo sapiens
NOV11k	175070808	57	58	Adiponectin like homo sapiens
NOV11l	175070812	59	60	Adiponectin like homo sapiens
NOV11m	175070828	61	62	Adiponectin like homo sapiens
NOV11n	175070836	63	64	Adiponectin like homo sapiens
NOV11o	175070840	65	66	Adiponectin like homo sapiens
NOV12a	CG95794-01	67	68	Trypsin III, cationic precursor like homo sapiens
NOV13a	CG95804-01	69	70	Tissue kallikrein like homo sapiens
NOV14a	CG95861-01	71	72	Transforming growth factor, beta-induced, 68kD
NOV15a	CG96412-01	73	74	Diphthamide synthesis protein
NOV15b	CG96412-03	75	76	Diphthamide synthesis protein
NOV15c	228116438	77	78	Diphthamide synthesis protein
NOV15d	228116442	79	80	Diphthamide synthesis protein
NOV16a	CG96511-01	81	82	Human WECHE Lungkine

NOV17a	CG96522-01	83	84	ADAM TS7
NOV18a	CG96535-01	85	86	Inactive palmitoyl-protein thioesterase-2I like homo sapiens
NOV18b	CG96535-02	87	88	Inactive palmitoyl-protein thioesterase-2I like homo sapiens
NOV19a	CG96567-02	89	90	Betacellulin precursor
NOV20a	CG96637-01	91	92	Small inducible cytokine A23 precursor
NOV20b	CG96637-04	93	94	Small inducible cytokine A23 precursor
NOV21a	CG97274-01	95	96	Granulocyte colony-stimulating factor precursor like homo sapiens
NOV21b	CG97274-03	97	98	Granulocyte colony-stimulating factor precursor like homo sapiens
NOV21c	CG97274-04	99	100	Granulocyte colony-stimulating factor precursor like homo sapiens
NOV21d	197208289	101	102	Granulocyte colony-stimulating factor precursor like homo sapiens
NOV22a	CG97288-01	103	104	Human platelet basic protein 2 like homo sapiens
NOV22b	CG97288-02	105	106	Human platelet basic protein 2 like homo sapiens
NOV23a	CG97516-01	107	108	Brain natriuretic peptide precursor like homo sapiens
NOV24a	CG97550-01	109	110	Serine protease like homo sapiens
NOV25a	CG97738-01	111	112	Acyl-CoA-binding protein (Diazepam binding inhibitor) like homo sapiens
NOV26a	CG97800-01	113	114	Elastase IV like homo sapiens
NOV26b	CG97800-02	115	116	Elastase IV like homo sapiens
NOV26c	CG97800-03	117	118	Elastase IV like homo sapiens
NOV27a	CG98092-01	119	120	Collagen like homo sapiens
NOV28a	CG98121-01	121	122	Viral receptor like homo sapiens
NOV29a	CG99662-01	123	124	Cathepsin L2 precursor

Table 1 indicates homology of NOVX nucleic acids to known protein families. Thus, the nucleic acids and polypeptides, antibodies and related compounds according to the invention corresponding to a NOVX as identified in column 1 of Table 1 will be useful in therapeutic and diagnostic applications implicated in, for example, pathologies and disorders associated with the known protein families identified in column 5 of Table 1.

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

Consistent with other known members of the family of proteins, identified in column 5 of Table 1, the NOVX polypeptides of the present invention show homology to, and contain domains that are characteristic of, other members of such protein families.

Details of the sequence relatedness and domain analysis for each NOVX are presented in Example A.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic
5   acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit diseases associated with the protein families listed in Table 1.

The NOVX nucleic acids and polypeptides are also useful for detecting specific cell types. Details of the expression analysis for each NOVX are presented in Example C.
10 Accordingly, the NOVX nucleic acids, polypeptides, antibodies and related compounds according to the invention will have diagnostic and therapeutic applications in the detection of a variety of diseases with differential expression in normal vs. diseased tissues, e.g. a variety of cancers.

Additional utilities for NOVX nucleic acids and polypeptides according to the
15 invention are disclosed herein.

NOVX clones

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the
20 presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

The NOVX genes and their corresponding encoded proteins are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy.
25 Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the NOVX genes, based on the tissues in which they are most highly expressed. Uses include developing products for the diagnosis or treatment of a variety of diseases and disorders.

30 The NOVX nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein

the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene

5   delivery/gene ablation), and (v) a composition promoting tissue regeneration *in vitro* and *in vivo* (vi) biological defense weapon.

In one specific embodiment, the invention includes an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO:2n,

10   wherein n is an integer between 1 and 62; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) an amino acid sequence selected from the group

15   consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; and (e) a fragment of any of (a) through (d).

20   In another specific embodiment, the invention includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 62; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID

25   NO:2n, wherein n is an integer between 1 and 62 wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62; (d) a variant of the amino acid sequence selected from

30   the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; (e) a nucleic acid fragment encoding at least a portion of a polypeptide

13

comprising the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 62 or any variant of said polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and (f) the

5   complement of any of said nucleic acid molecules.

In yet another specific embodiment, the invention includes an isolated nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of: (a) the nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 62; (b) a nucleotide

10 sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 62 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; (c) a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID

15 NO:2n-1, wherein n is an integer between 1 and 62; and (d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 62 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

20

NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify

25 NOVX-encoding nucleic acids (e.g., NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA), RNA molecules (e.g., mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The

30 nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

A NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a "mature" form of a polypeptide or protein disclosed in the present invention is the product

14

of a naturally occurring polypeptide, precursor form, or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein.

5   The product "mature" form arises, by way of nonlimiting example, as a result of one or more naturally occurring processing steps that may take place within the cell (host cell) in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF or the proteolytic cleavage of a signal peptide

10   or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is

15   cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the

20   operation of only one of these processes, or a combination of any of them.

The term "probe", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), and 100 nt, or as many as approximately, e.g., 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length

25   probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as used herein, is a nucleic acid which

30   is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived.

For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, 0.1 kb, or less of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, etc.).

5   Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, culture medium, or of chemical precursors or other chemicals.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 62, or a

10   complement of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A

15   LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993).

A nucleic acid of the invention can be amplified using cDNA, mRNA or, alternatively, genomic DNA as a template with appropriate oligonucleotide primers

20   according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide

25   residues. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide

30   comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment

5   encoding a biologically-active portion of A NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence shown SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, is one that is sufficiently complementary to the nucleotide sequence shown SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62,that it can hydrogen bond with few or no mismatches to the nucleotide sequence shown SEQ ID

10   NOS:2n-1, wherein n is an integer between 1 and 62, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic,

15   non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

20   "Fragments" provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid

25   sequence of choice.

A full-length NOVX clone is identified as containing an ATG translation start codon and an in-frame stop codon. Any disclosed NOVX nucleotide sequence lacking an ATG start codon therefore encodes a truncated C-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 5'

30   direction of the disclosed sequence. Any disclosed NOVX nucleotide sequence lacking an in-frame stop codon similarly encodes a truncated N-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 3' direction of the disclosed sequence.

"Derivatives" are nucleic acid sequences or amino acid sequences formed from the native compounds either directly, by modification, or by partial substitution. "Analogs" are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound, e.g. they differ from it in respect to certain components

5   or side chains. Analogs may be synthetic or derived from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

Derivatives and analogs may be full length or other than full length. Derivatives or

10   analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done

15   by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the proteins of the invention under stringent, moderately stringent, or low stringent conditions. *See e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

20   A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences include those sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of

25   RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for A NOVX polypeptide of species other than humans, including, but not limited to vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring

30   allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding a human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see

18

below) in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

A NOVX polypeptide is encoded by the open reading frame ("ORF") of a NOVX
5   nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be
10   any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, *e.g.,* a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human NOVX genes
15   allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.* from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises a substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150,
20   200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62; or an anti-sense strand nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62; or of a naturally occurring mutant of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62.

Probes based on the human NOVX nucleotide sequences can be used to detect
25   transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe has a detectable label attached, *e.g.* the label can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express A NOVX protein, such as by measuring a level of A NOVX-encoding nucleic
30   acid in a sample of cells from a subject *e.g.,* detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of A NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical, an activity

19

of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, that encodes a polypeptide having A

5   NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

NOVX Nucleic Acid and Polypeptide Variants

10   The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62. In another embodiment, an isolated nucleic acid molecule of

15   the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NOS:2n, wherein n is an integer between 1 and 62.

In addition to the human NOVX nucleotide sequences shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid

20   sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding A NOVX protein, preferably a vertebrate NOVX protein.

25   Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX polypeptides, are intended to be within the scope of the invention.

30   Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from the human SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and

homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

5   Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, 2000 or more nucleotides in length. In yet another

10   embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least about 65% homologous to each other typically remain hybridized to each other.

          Homologs (i.e., nucleic acids encoding NOVX proteins derived from species other

15   than human) or other related sequences (e.g., paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

          As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to

20   no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid

25   concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at

30   pH 7.0 to 8.3 and the temperature is at least about 30 °C for short probes, primers or oligonucleotides (e.g., 10 nt to 50 nt) and at least about 60 °C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in

Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley &

Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least

about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically

5   remain hybridized to each other. A non-limiting example of stringent hybridization

conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl

(pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured

salmon sperm DNA at 65 °C, followed by one or more washes in 0.2X SSC, 0.01% BSA

at 50 °C. An isolated nucleic acid molecule of the invention that hybridizes under

10   stringent conditions to the sequences SEQ ID NOS:2n-1, wherein n is an integer between

1 and 62, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a

"naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a

nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic

15   acid molecule comprising the nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an

integer between 1 and 62, or fragments, analogs or derivatives thereof, under conditions

of moderate stringency is provided. A non-limiting example of moderate stringency

hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS

and 100 mg/ml denatured salmon sperm DNA at 55 °C, followed by one or more washes

20   in 1X SSC, 0.1% SDS at 37 °C. Other conditions of moderate stringency that may be used

are well-known within the art. *See, e.g.*, Ausubel, et *al.* (eds.), 1993, CURRENT PROTOCOLS

IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990; GENE TRANSFER

AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid

25   molecule comprising the nucleotide sequences SEQ ID NOS:2n-1, wherein n is an integer

between 1 and 62, or fragments, analogs or derivatives thereof, under conditions of low

stringency, is provided. A non-limiting example of low stringency hybridization

conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5

mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm

30   DNA, 10% (wt/vol) dextran sulfate at 40 °C, followed by one or more washes in 2X SSC,

25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50 °C. Other conditions of low

stringency that may be used are well known in the art (*e.g.*, as employed for cross-species

hybridizations). *See, e.g.,* Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN

MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND
EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981.
Proc Natl Acad Sci USA **78**: 6789-6792.

5   **Conservative Mutations**

In addition to naturally-occurring allelic variants of NOVX sequences that may
exist in the population, the skilled artisan will further appreciate that changes can be
introduced by mutation into the nucleotide sequences SEQ ID NOS:2n-1, wherein n is an
integer between 1 and 62, thereby leading to changes in the amino acid sequences of the
10   encoded NOVX proteins, without altering the functional ability of the NOVX proteins.
For example, nucleotide substitutions leading to amino acid substitutions at
"non-essential" amino acid residues can be made in the sequence SEQ ID NOS:2n,
wherein n is an integer between 1 and 62. A "non-essential" amino acid residue is a
residue that can be altered from the wild-type sequences of the NOVX proteins without
15   altering their biological activity, whereas an "essential" amino acid residue is required for
such biological activity. For example, amino acid residues that are conserved among the
NOVX proteins of the invention are predicted to be particularly non-amenable to
alteration. Amino acids for which conservative substitutions can be made are well known
within the art.

20   Another aspect of the invention pertains to nucleic acid molecules encoding
NOVX proteins that contain changes in amino acid residues that are not essential for
activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NOS:2n-1,
wherein n is an integer between 1 and 62, yet retain biological activity. In one
embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding
25   a protein, wherein the protein comprises an amino acid sequence at least about 50%
homologous to the amino acid sequences SEQ ID NOS:2n, wherein n is an integer
between 1 and 62. Preferably, the protein encoded by the nucleic acid molecule is at least
about 60% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 62;
more preferably at least about 70% homologous SEQ ID NOS:2n, wherein n is an integer
30   between 1 and 62; still more preferably at least about 80% homologous to SEQ ID
NOS:2n, wherein n is an integer between 1 and 62; even more preferably at least about
90% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 62; and most

preferably at least about 95% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 62.

An isolated nucleic acid molecule encoding A NOVX protein homologous to the protein of SEQ ID NOS:2n, wherein n is an integer between 1 and 62, can be created by
5   introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NOS:2n-1, wherein n is an integer
10   between 1 and 62, by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains
15   have been defined within the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine,
20   isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of A NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for
25   NOVX biological activity to identify mutants that retain activity. Following mutagenesis SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain
30   interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids

that may be substituted for each other. Likewise, the "weak" group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, HFY, wherein the letters within each group represent the single letter amino acid code.

5   In one embodiment, a mutant NOVX protein can be assayed for (*i*) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or biologically-active portions thereof, (*ii*) complex formation between a mutant NOVX protein and A NOVX ligand; or (*iii*) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically-active portion thereof; (*e.g.* avidin proteins).

10   In yet another embodiment, a mutant NOVX protein can be assayed for the ability to regulate a specific biological function (*e.g.*, regulation of insulin release).

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid

15   molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*, complementary to the coding strand of a double-stranded cDNA molecule

20   or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of A NOVX protein of SEQ ID NOS:2n, wherein n is an integer between 1 and 62, or

25   antisense nucleic acids complementary to A NOVX nucleic acid sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding A NOVX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons,

30   which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to

5' and 3' sequences, which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, beta-D-mannosylqueosine, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed

from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding A NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. A α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other. *See, e.g.*, Gaultier, *et al.*, 1987. *Nucl. Acids Res.* **15**: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (*See, e.g.*, Inoue, *et al.* 1987. *Nucl. Acids Res.* **15**: 6131-6148) or a chimeric RNA-DNA analogue (*See, e.g.*, Inoue, *et al.*, 1987. *FEBS Lett.* **215**: 327-330.

Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the 5 modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a 10 complementary region. Thus, ribozymes (e.g., hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of A NOVX cDNA disclosed herein (*i.e.*, SEQ ID NOS:2n-1, 15 wherein n is an integer between 1 and 62). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a NOVX-encoding mRNA. *See, e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease 20 activity from a pool of RNA molecules. *See, e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent 25 transcription of the NOVX gene in target cells. *See, e.g.*, Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, 30 or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. *See, e.g.*, Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose

phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis

5　　protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or

10　　inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S_1 nucleases (*See,* Hyrup, *et al.*, 1996.*supra*); or as probes or primers for DNA sequence and hybridization (*See,* Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.*, 1996. *supra*).

15　　In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow

20　　DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (*see,* Hyrup, *et al.*, 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in

25　　Hyrup, *et al.*, 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.*, Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988.

30　　PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.*, Finn, *et al.*, 1996. *supra.* Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.*, Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT

5   Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g.*, Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered

10   cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

NOVX Polypeptides

A polypeptide according to the invention includes a polypeptide including the

15   amino acid sequence of NOVX polypeptides whose sequences are provided in SEQ ID NOS:2n, wherein n is an integer between 1 and 62. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in SEQ ID NOS:2n, wherein n is an integer between 1 and 62, while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional

20   fragment thereof.

In general, A NOVX variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting

25   one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs

30   thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by

recombinant DNA techniques. Alternative to recombinant expression, A NOVX protein
or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion
thereof is substantially free of cellular material or other contaminating proteins from the
5   cell or tissue source from which the NOVX protein is derived, or substantially free from
chemical precursors or other chemicals when chemically synthesized. The language
"substantially free of cellular material" includes preparations of NOVX proteins in which
the protein is separated from cellular components of the cells from which it is isolated or
recombinantly-produced. In one embodiment, the language "substantially free of cellular
10   material" includes preparations of NOVX proteins having less than about 30% (by dry
weight) of non-NOVX proteins (also referred to herein as a "contaminating protein"),
more preferably less than about 20% of non-NOVX proteins, still more preferably less
than about 10% of non-NOVX proteins, and most preferably less than about 5% of
non-NOVX proteins. When the NOVX protein or biologically-active portion thereof is
15   recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*,
culture medium represents less than about 20%, more preferably less than about 10%, and
most preferably less than about 5% of the volume of the NOVX protein preparation.

The language "substantially free of chemical precursors or other chemicals"
includes preparations of NOVX proteins in which the protein is separated from chemical
20   precursors or other chemicals that are involved in the synthesis of the protein. In one
embodiment, the language "substantially free of chemical precursors or other chemicals"
includes preparations of NOVX proteins having less than about 30% (by dry weight) of
chemical precursors or non-NOVX chemicals, more preferably less than about 20%
chemical precursors or non-NOVX chemicals, still more preferably less than about 10%
25   chemical precursors or non-NOVX chemicals, and most preferably less than about 5%
chemical precursors or non-NOVX chemicals.

Biologically-active portions of NOVX proteins include peptides comprising amino
acid sequences sufficiently homologous to or derived from the amino acid sequences of
the NOVX proteins (*e.g.*, the amino acid sequence shown in SEQ ID NOS:2n, wherein n is
30   an integer between 1 and 62) that include fewer amino acids than the full-length NOVX
proteins, and exhibit at least one activity of A NOVX protein. Typically,
biologically-active portions comprise a domain or motif with at least one activity of the

31

NOVX protein. A biologically-active portion of A NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of

5   the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence shown SEQ ID NOS:2n, wherein n is an integer between 1 and 62. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 62, and retains the functional activity of the protein of SEQ ID NOS:2n, wherein n is

10   an integer between 1 and 62, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence SEQ ID NOS:2n, wherein n is an integer between 1 and 62, and retains the functional activity of the NOVX proteins of

15   SEQ ID NOS:2n, wherein n is an integer between 1 and 62.

DETERMINING HOMOLOGY BETWEEN TWO OR MORE SEQUENCES

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be

20   introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the

25   molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. See,

30   Needleman and Wunsch, 1970. J Mol Biol 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%,

80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region

5   of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of

10   comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent

15   sequence identity as compared to a reference sequence over a comparison region.

CHIMERIC AND FUSION PROTEINS

The invention also provides NOVX chimeric or fusion proteins. As used herein, A NOVX "chimeric protein" or "fusion protein" comprises A NOVX polypeptide

20   operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to A NOVX protein SEQ ID NOS:2n, wherein n is an integer between 1 and 62, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from

25   the NOVX protein and that is derived from the same or a different organism. Within A NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of A NOVX protein. In one embodiment, A NOVX fusion protein comprises at least one biologically active portion of A NOVX protein. In another embodiment, A NOVX fusion protein comprises at least two biologically active portions of A NOVX protein. In yet

30   another embodiment, A NOVX fusion protein comprises at least three biologically active portions of A NOVX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused

33

in-frame with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase)

5   sequences. Such fusion proteins can facilitate the purification of recombinant NOVX polypeptides.

In another embodiment, the fusion protein is A NOVX protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of NOVX can be increased through use of a

10   heterologous signal sequence.

In yet another embodiment, the fusion protein is a NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a

15   subject to inhibit an interaction between A NOVX ligand and A NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of A NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well

20   as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with A NOVX ligand.

A NOVX chimeric or fusion protein of the invention can be produced by standard

25   recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic

30   ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and

reamplified to generate a chimeric gene sequence (*see, e.g.,* Ausubel, *et al.* (eds.)
CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover,
many expression vectors are commercially available that already encode a fusion moiety
(*e.g.,* a GST polypeptide). A NOVX-encoding nucleic acid can be cloned into such an

5   expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

NOVX AGONISTS AND ANTAGONISTS

The invention also pertains to variants of the NOVX proteins that function as either
NOVX agonists (*i.e.,* mimetics) or as NOVX antagonists. Variants of the NOVX protein

10   can be generated by mutagenesis (*e.g.,* discrete point mutation or truncation of the NOVX
protein). An agonist of the NOVX protein can retain substantially the same, or a subset of,
the biological activities of the naturally occurring form of the NOVX protein. An
antagonist of the NOVX protein can inhibit one or more of the activities of the naturally
occurring form of the NOVX protein by, for example, competitively binding to a

15   downstream or upstream member of a cellular signaling cascade, which includes the
NOVX protein. Thus, specific biological effects can be elicited by treatment with a
variant of limited function. In one embodiment, treatment of a subject with a variant
having a subset of the biological activities of the naturally occurring form of the protein
has fewer side effects in a subject relative to treatment with the naturally occurring form of

20   the NOVX proteins.

Variants of the NOVX proteins that function as either NOVX agonists (*i.e.,*
mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries
of mutants (*e.g.,* truncation mutants) of the NOVX proteins for NOVX protein agonist or
antagonist activity. In one embodiment, a variegated library of NOVX variants is

25   generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a
variegated gene library. A variegated library of NOVX variants can be produced by, for
example, enzymatically ligating a mixture of synthetic oligonucleotides into gene
sequences such that a degenerate set of potential NOVX sequences is expressible as
individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.,* for phage

30   display) containing the set of NOVX sequences therein. There are a variety of methods,
which can be used to produce libraries of potential NOVX variants from a degenerate
oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be
performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an

appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are well known within the art. *See, e.g.,* Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.,* 1984. *Annu. Rev.*
5   *Biochem.* 53: 323; Itakura, *et al.,* 1984. *Science* 198: 1056; Ike, *et al.,* 1983. *Nucl. Acids Res.* 11: 477.

POLYPEPTIDE LIBRARIES

In addition, libraries of fragments of the NOVX protein coding sequences can be
10   used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of A NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of A NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA
15   to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S_1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX proteins.

20   Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high
25   throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that
30   enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants. *See, e.g.,* Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.,* 1993. *Protein Engineering* 6:327-331.

NOVX Antibodies

The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that

5   contain an antigen-binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, antibody molecules obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule.

10   Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated protein of the invention intended to serve as an antigen, or a portion or

15   fragment thereof, can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid

20   sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NOs: 2n, wherein n is an integer between 1 and 62, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid

25   residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX that is located on the surface of the protein, *e.g.*, a

30   hydrophilic region. A hydrophobicity analysis of the human NOVX protein sequence will indicate which regions of a NOVX polypeptide are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and

hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each incorporated herein

5   by reference in their entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that

10   immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor

15   Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (e.g.,

20   rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore,

25   the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete

30   and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants which

can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell [Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103]. Immortalized cell lines are usually transformed mammalian

39

cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells

5   lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

        Preferred immortalized cell lines are those that fuse efficiently, support stable high

10   level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also

15   have been described for the production of human monoclonal antibodies [Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63].

        The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the

20   binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and

25   Pollard, Anal. Biochem., 107:220 (1980). It is an objective, especially important in therapeutic applications of monoclonal antibodies, to identify antibodies having a high degree of specificity and a high binding affinity for the target antigen.

        After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods (Goding,1986). Suitable

30   culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

5   The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells

10   of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the

15   coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or

20   can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further

25   comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')$_2$ or other antigen-binding subsequences of antibodies) that are principally comprised

30   of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by

substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the

5   recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody

10   optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

Human Antibodies

15   Fully human antibodies essentially relate to antibody molecules in which the entire sequence of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol

20   Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with

25   Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by

30   introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire.

This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al,(Nature Biotechnology 14, 845-51 (1996)); Neuberger
5   (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT
10   publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides
15   all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can
20   be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to
25   obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to
30   prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem

cell a transgenic mouse whose somatic and germ cells contain the gene encoding the
selectable marker.

A method for producing an antibody of interest, such as a human antibody, is
disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that
5   contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in
culture, introducing an expression vector containing a nucleotide sequence encoding a
light chain into another mammalian host cell, and fusing the two cells to form a hybrid
cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically
10   relevant epitope on an immunogen, and a correlative method for selecting an antibody that
binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT
publication WO 99/53049.

F_{ab} Fragments and Single Chain Antibodies

15   According to the invention, techniques can be adapted for the production of
single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S.
Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab}
expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid
and effective identification of monoclonal F_{ab} fragments with the desired specificity for a
20   protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that
contain the idiotypes to a protein antigen may be produced by techniques known in the art
including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an
antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an
$F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule
25   with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies
that have binding specificities for at least two different antigens. In the present case, one
30   of the binding specificities is for an antigenic protein of the invention. The second binding
target is any other antigen, and advantageously is a cell-surface protein or receptor or
receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random

5   assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659

10   (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to

15   have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh

20   et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or

25   more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the

30   heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. F(ab')₂ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example,

45

bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science
229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to
generate F(ab')$_2$ fragments. These fragments are reduced in the presence of the dithiol
complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular
5   disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate
(TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol
by reduction with mercaptoethylamine and is mixed with an equimolar amount of the
other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies
produced can be used as agents for the selective immobilization of enzymes.

10   Additionally, Fab' fragments can be directly recovered from E. coli and chemically
coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992)
describe the production of a fully humanized bispecific antibody F(ab')$_2$ molecule. Each
Fab' fragment was separately secreted from E. coli and subjected to directed chemical
coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was
15   able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well
as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor
targets.

          Various techniques for making and isolating bispecific antibody fragments
directly from recombinant cell culture have also been described. For example, bispecific
20   antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol.
148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were
linked to the Fab' portions of two different antibodies by gene fusion. The antibody
homodimers were reduced at the hinge region to form monomers and then re-oxidized to
form the antibody heterodimers. This method can also be utilized for the production of
25   antibody homodimers. The "diabody" technology described by Hollinger et al., Proc.
Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for
making bispecific antibody fragments. The fragments comprise a heavy-chain variable
domain (V$_H$) connected to a light-chain variable domain (V$_L$) by a linker which is too short
to allow pairing between the two domains on the same chain. Accordingly, the V$_H$ and V$_L$
30   domains of one fragment are forced to pair with the complementary V$_L$ and V$_H$ domains of
another fragment, thereby forming two antigen-binding sites. Another strategy for making
bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been
reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC).

See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has

5   dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody

10 conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used

15 include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the

20 tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as

25 dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as

30 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an

exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

Immunoliposomes

The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., Proc. Natl. Acad. Sci. USA, 82: 3688 (1985); Hwang et al., Proc. Natl Acad. Sci. USA, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in Martin et al ., J. Biol. Chem., 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome. See Gabizon et al., J. National Cancer Inst., 81(19): 1484 (1989).

Diagnostic Applications of Antibodies Directed Against the Proteins of the Invention

Antibodies directed against a protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of the protein (e.g., for use in measuring levels of the protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies against the proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antigen binding domain, are utilized as pharmacologically-active compounds (see below).

An antibody specific for a protein of the invention can be used to isolate the protein by standard techniques, such as immunoaffinity chromatography or

immunoprecipitation. Such an antibody can facilitate the purification of the natural protein antigen from cells and of recombinantly produced antigen expressed in host cells. Moreover, such an antibody can be used to detect the antigenic protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of

5   the antigenic protein. Antibodies directed against the protein can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials,

10   luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine

15   fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

        Antibody Therapeutics

20   Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may used as therapeutic agents. Such agents will generally be employed to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the

25   target. Such an effect may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule and the target antigen in question. In the first instance, administration of the antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the

30   ligand serves as an effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

               Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case

the target, a receptor having an endogenous ligand which may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

A therapeutically effective amount of an antibody of the invention relates generally

5   to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which

10   an administered antibody is depleted from the free volume other subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

15

Pharmaceutical Compositions of Antibodies

Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and

20   considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed. (Alfonso R. Gennaro, et al., editors) Mack Pub. Co., Easton, Pa. : 1995; Drug Absorption Enhancement : Concepts, Possibilities, Limitations, And Trends, Harwood Academic Publishers, Langhorne, Pa., 1994; and Peptide And Protein Drug

25   Delivery (Advances In Parenteral Sciences, Vol. 4), 1991, M. Dekker, New York.

If the antigenic protein is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment that specifically binds to the binding domain of the target

30   protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology. See, e.g., Marasco et al., Proc. Natl. Acad. Sci. USA, 90:

7889-7893 (1993). The formulation herein can also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition can comprise an agent that enhances its function, such as, for

5   example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended.

The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example,

10   hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

The formulations to be used for in vivo administration must be sterile. This is

15   readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, e.g., films, or microcapsules. Examples of sustained-release matrices include polyesters,

20   hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and γ ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT ™ (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and

25   poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

ELISA Assay

30   An agent for detecting an analyte protein is an antibody capable of binding to an analyte protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., F_{ab} or $F_{(ab)2}$) can be used. The term "labeled", with regard to the probe or antibody,

is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a

5   fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. Included within the usage of the term "biological sample", therefore, is blood and a fraction or component of blood

10   including blood serum, blood plasma, or lymph. That is, the detection method of the invention can be used to detect an analyte mRNA, protein, or genomic DNA in a biological sample in vitro as well as in vivo. For example, in vitro techniques for detection of an analyte mRNA include Northern hybridizations and in situ hybridizations. In vitro techniques for detection of an analyte protein include enzyme linked immunosorbent

15   assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. In vitro techniques for detection of an analyte genomic DNA include Southern hybridizations. Procedures for conducting immunoassays are described, for example in "ELISA: Theory and Practice: Methods in Molecular Biology", Vol. 42, J. R. Crowther (Ed.) Human Press, Totowa, NJ, 1995; "Immunoassay", E. Diamandis and T. Christopoulus, Academic Press,

20   Inc., San Diego, CA, 1996; and "Practice and Thory of Enzyme Immunoassays", P. Tijssen, Elsevier Science Publishers, Amsterdam, 1985. Furthermore, in vivo techniques for detection of an analyte protein include introducing into a subject a labeled anti-an analyte protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging

25   techniques.

NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding A NOVX protein, or derivatives, fragments, analogs or

30   homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein

additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the

5 genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and

10 "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the

15 invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to

20 the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory

25 sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by

30 those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic

54

acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX

5   proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example

10   using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such

15   fusion vectors typically serve three purposes: (*i*) to increase expression of recombinant protein; (*ii*) to increase the solubility of the recombinant protein; and (*iii*) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant

20   protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase

25   (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990)

30   60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. *See, e.g.,* Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN

ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy
is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression
vector so that the individual codons for each amino acid are those preferentially utilized in
E. coli (*see, e.g.,* Wada, *et al.,* 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of
5   nucleic acid sequences of the invention can be carried out by standard DNA synthesis
techniques.

In another embodiment, the NOVX expression vector is a yeast expression vector.
Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1
(Baldari, *et al.,* 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell*
10   30: 933-943), pJRY88 (Schultz *et al.,* 1987. *Gene* 54: 113-123), pYES2 (Invitrogen
Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, NOVX can be expressed in insect cells using baculovirus expression
vectors. Baculovirus vectors available for expression of proteins in cultured insect cells
(*e.g.,* SF9 cells) include the pAc series (Smith, *et al.,* 1983. *Mol. Cell. Biol.* 3: 2156-2165)
15   and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in
mammalian cells using a mammalian expression vector. Examples of mammalian
expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC
(Kaufman, *et al.,* 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the
20   expression vector's control functions are often provided by viral regulatory elements. For
example, commonly used promoters are derived from polyoma, adenovirus 2,
cytomegalovirus, and simian virus 40. For other suitable expression systems for both
prokaryotic and eukaryotic cells see, *e.g.,* Chapters 16 and 17 of Sambrook, *et al.,*
MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor
25   Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable
of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.,*
tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific
regulatory elements are known in the art. Non-limiting examples of suitable
30   tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.,*
1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988.
Adv. Immunol. 43: 235-275), in particular promoters of T cell receptors (Winoto and
Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.,* 1983. *Cell*

33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters
(*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86:
5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and
mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316

5   and European Application Publication No. 264,166). Developmentally-regulated
promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990.
Science 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989.
Genes Dev. 3: 537-546).

           The invention further provides a recombinant expression vector comprising a DNA

10   molecule of the invention cloned into the expression vector in an antisense orientation.
That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that
allows for expression (by transcription of the DNA molecule) of an RNA molecule that is
antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid
cloned in the antisense orientation can be chosen that direct the continuous expression of

15   the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or
enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or
cell type specific expression of antisense RNA. The antisense expression vector can be in
the form of a recombinant plasmid, phagemid or attenuated virus in which antisense
nucleic acids are produced under the control of a high efficiency regulatory region, the

20   activity of which can be determined by the cell type into which the vector is introduced.
For a discussion of the regulation of gene expression using antisense genes *see, e.g.,*
Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis,"
Reviews-Trends in Genetics, Vol. 1(1) 1986.

           Another aspect of the invention pertains to host cells into which a recombinant

25   expression vector of the invention has been introduced. The terms "host cell" and
"recombinant host cell" are used interchangeably herein. It is understood that such terms
refer not only to the particular subject cell but also to the progeny or potential progeny of
such a cell. Because certain modifications may occur in succeeding generations due to
either mutation or environmental influences, such progeny may not, in fact, be identical to

30   the parent cell, but are still included within the scope of the term as used herein.

           A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein
can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells

(such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms

5 "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (MOLECULAR CLONING: A

10 LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these

15 integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector.

20 Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) NOVX protein. Accordingly, the invention further

25 provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

30

Transgenic NOVX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized

oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been

5   altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates,

10   sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a

15   mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing

20   NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (e.g., by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human NOVX cDNA sequences SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human

25   NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further supra) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of NOVX protein to

30   particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory

Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed

5   additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of A NOVX gene into which a deletion, addition or substitution has been

10   introduced to thereby alter, e.g., functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (e.g., the cDNA of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, can be used to construct a homologous

15   recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous

20   recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to

25   occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. See, e.g., Thomas, et al., 1987. Cell 51: 503 for a description of homologous

30   recombination vectors. The vector is ten introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. See, e.g., Li, et al., 1992. Cell 69: 915.

60

The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female

5   foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991.

10   *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a

15   description of the cre/loxP recombinase system, *See, e.g.*, Lakso, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae*. *See,* O'Gorman, *et al.*, 1991. *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a

20   selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced

25   according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such

30   that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

Pharmaceutical Compositions

The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™

5   (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol,

10   polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for

15   example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

20   Sterile injectable solutions can be prepared by incorporating the active compound (e.g., A NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required

25   other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

            Oral compositions generally include an inert diluent or an edible carrier. They can

30   be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is

63

applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to
5   produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

10   The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.,* U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.,* Chen, *et al.,* 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the
15   gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.,* retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

20   The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods
          The isolated nucleic acid molecules of the invention can be used to express NOVX
25   protein (*e.g.,* via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (*e.g.,* in a biological sample) or a genetic lesion in A NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or
30   excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein (*e.g.*; diabetes (regulates insulin release); obesity (binds and transport lipids); metabolic disturbances associated with obesity, the metabolic syndrome X as well as anorexia and wasting

disorders associated with chronic diseases and various cancers, and infectious disease(possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. In yet a further aspect, the invention can be used in methods

5   to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

10   **Screening Assays**

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein

15   activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of A NOVX protein or polypeptide or biologically-active portion thereof. The test compounds

20   of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is

25   limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.,* Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small

30   molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the

art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*,

1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.*

37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed.*

5   *Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et*

al., 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992.

Biotechniques 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips

(Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores

10   (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA*

89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990.

Science 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382;

Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses

15   a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the

cell surface is contacted with a test compound and the ability of the test compound to bind

to A NOVX protein determined. The cell, for example, can of mammalian origin or a

yeast cell. Determining the ability of the test compound to bind to the NOVX protein can

be accomplished, for example, by coupling the test compound with a radioisotope or

20   enzymatic label such that binding of the test compound to the NOVX protein or

biologically-active portion thereof can be determined by detecting the labeled compound

in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either

directly or indirectly, and the radioisotope detected by direct counting of radioemission or

by scintillation counting. Alternatively, test compounds can be enzymatically-labeled

25   with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the

enzymatic label detected by determination of conversion of an appropriate substrate to

product. In one embodiment, the assay comprises contacting a cell which expresses a

membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the

cell surface with a known compound which binds NOVX to form an assay mixture,

30   contacting the assay mixture with a test compound, and determining the ability of the test

compound to interact with A NOVX protein, wherein determining the ability of the test

compound to interact with A NOVX protein comprises determining the ability of the test

compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion

5   thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or

10   interact with A NOVX target molecule. As used herein, a "target molecule" is a molecule with which A NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses A NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule

15   can be a non-NOVX molecule or A NOVX protein or polypeptide of the invention. In one embodiment, A NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (e.g. a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic

20   activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with A NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX

25   protein to bind to or interact with A NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (i.e. intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene

30   (comprising A NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, e.g., luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting A NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein

5   can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with A NOVX protein, wherein determining the ability of the test

10   compound to interact with A NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and

15   determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to A NOVX target molecule by one of the methods described above for determining direct binding. In an

20   alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate A NOVX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

25   In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with A NOVX protein, wherein determining the ability of the test compound to interact with A NOVX protein comprises

30   determining the ability of the NOVX protein to preferentially bind to or modulate the activity of A NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising

the membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing
agent such that the membrane-bound form of NOVX protein is maintained in solution.
Examples of such solubilizing agents include non-ionic detergents such as
n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide,
5   decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®,
Isotridecypoly(ethylene glycol ether)$_n$, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane
sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or
3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may
10   be desirable to immobilize either NOVX protein or its target molecule to facilitate
separation of complexed from uncomplexed forms of one or both of the proteins, as well
as to accommodate automation of the assay. Binding of a test compound to NOVX
protein, or interaction of NOVX protein with a target molecule in the presence and
absence of a candidate compound, can be accomplished in any vessel suitable for
15   containing the reactants. Examples of such vessels include microtiter plates, test tubes,
and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that
adds a domain that allows one or both of the proteins to be bound to a matrix. For
example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto
glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized
20   microtiter plates, that are then combined with the test compound or the test compound and
either the non-adsorbed target protein or NOVX protein, and the mixture is incubated
under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt
and pH). Following incubation, the beads or microtiter plate wells are washed to remove
any unbound components, the matrix immobilized in the case of beads, complex
25   determined either directly or indirectly, for example, as described, *supra*. Alternatively,
the complexes can be dissociated from the matrix, and the level of NOVX protein binding
or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the
screening assays of the invention. For example, either the NOVX protein or its target
30   molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated
NOVX protein or target molecules can be prepared from biotin-NHS
(N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation
kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated

70

96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such

5   complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in

10   a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein

15   expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (i.e., statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound

20   than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (see, e.g., U.S. Patent No. 5,283,317;

25   Zervos, et al., 1993. Cell 72: 223-232; Madura, et al., 1993. J. Biol. Chem. 268: 12046-12054; Bartel, et al., 1993. Biotechniques 14: 920-924; Iwabuchi, et al., 1993. Oncogene 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also likely to be involved in the propagation of

30   signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the

71

assay utilizes two different DNA constructs. In one construct, the gene that codes for

NOVX is fused to a gene encoding the DNA binding domain of a known transcription

factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA

sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that

5   codes for the activation domain of the known transcription factor. If the "bait" and the

"prey" proteins are able to interact, *in vivo*, forming A NOVX-dependent complex, the

DNA-binding and activation domains of the transcription factor are brought into close

proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is

operably linked to a transcriptional regulatory site responsive to the transcription factor.

10   Expression of the reporter gene can be detected and cell colonies containing the functional

transcription factor can be isolated and used to obtain the cloned gene that encodes the

protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned

screening assays and uses thereof for treatments as described herein.

15

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the

corresponding complete gene sequences) can be used in numerous ways as polynucleotide

reagents. By way of example, and not of limitation, these sequences can be used to: (*i*)

20   map their respective genes on a chromosome; and, thus, locate gene regions associated

with genetic disease; (*ii*) identify an individual from a minute biological sample (tissue

typing); and (*iii*) aid in forensic identification of a biological sample. Some of these

applications are described in the subsections, below.

25   ### Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this

sequence can be used to map the location of the gene on a chromosome. This process is

called chromosome mapping. Accordingly, portions or fragments of the NOVX

sequences, SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, or fragments or

30   derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a

chromosome. The mapping of the NOVX sequences to chromosomes is an important first

step in correlating these sequences with genes associated with disease.

Briefly, NOVX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the NOVX sequences. Computer analysis of the NOVX, sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers

5   can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the NOVX sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and mouse cells). As hybrids of human and mouse cells grow and divide,

10   they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a

15   single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. See, e.g., D'Eustachio, et al., 1983. Science 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

20   PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOVX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

25   Fluorescence in situ hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern

30   of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple

detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see*, Verma, *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

5   Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance

10   of cross hybridizations during chromosomal mapping.

         Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship

15   between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

         Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the NOVX gene, can be determined. If a

20   mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence.

25   Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

         The NOVX sequences of the invention can also be used to identify individuals

30   from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional

DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of

5   an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a

10   unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation

15   between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes.

20   Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62,

25   are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic

30   assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of

a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia,

5   cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of

10   developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in A NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

15   Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the

20   individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

              Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

              These and other agents are described in further detail in the following sections.

25

              DIAGNOSTIC ASSAYS

              An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX

30   protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length

NOVX nucleic acid, such as the nucleic acid of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 62, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in
5   the diagnostic assays of the invention are described herein.

An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')$_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended
10   to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with
15   fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample in vitro as well as in vivo. For example, in vitro techniques for detection of NOVX mRNA include
20   Northern hybridizations and in situ hybridizations. In vitro techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. In vitro techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, in vivo techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX
25   antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is
30   a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of

NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

PROGNOSTIC ASSAYS

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with

aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (e.g., wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

5   The methods of the invention can also be used to detect genetic lesions in A NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration

10   affecting the integrity of a gene encoding A NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (i) a deletion of one or more nucleotides from A NOVX gene; (ii) an addition of one or more nucleotides to A NOVX gene; (iii) a substitution of one or more nucleotides of A NOVX gene, (iv) a chromosomal rearrangement of A NOVX gene;

15   (v) an alteration in the level of a messenger RNA transcript of A NOVX gene, (vi) aberrant modification of A NOVX gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of A NOVX gene, (viii) a non-wild-type level of A NOVX protein, (ix) allelic loss of A NOVX gene, and (x) inappropriate post-translational modification of A NOVX protein. As

20   described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in A NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

25   In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran, et al., 1988. Science 241: 1077-1080; and Nakazawa, et al., 1994. Proc. Natl. Acad. Sci. USA 91: 360-364), the latter of which can

30   be particularly useful for detecting point mutations in the NOVX-gene (see, Abravaya, et al., 1995. Nucl. Acids Res. 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that

79

specifically hybridize to A NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be

5   desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see*, Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see*, Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177);

10   Qβ Replicase (*see*, Lizardi, *et al*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

15   In an alternative embodiment, mutations in A NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA

20   indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.,* U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g.,* DNA or RNA, to high-density arrays

25   containing hundreds or thousands of oligonucleotides probes. *See, e.g.,* Cronin, *et al.,* 1996. *Human Mutation* 7: 244-255; Kozal, *et al.,* 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al., supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a

30   sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays

complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the
5   art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing
10   procedures can be utilized when performing the diagnostic assays (*see, e.g.,* Naeve, *et al.,* 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, *e.g.,* PCT International Publication No. WO 94/16101; Cohen, *et al.,* 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.,* 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in
15   which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g.,* Myers, *et al.,* 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded
20   duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S_1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with
25   hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.,* Cotton, *et al.,* 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.,* 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be
30   labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations

in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.,* Hsu, *et al.,* 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on A NOVX sequence, *e.g.,* a

5   wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to

10   identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.,* Orita, *et al.,* 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control NOVX

15   nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is

20   more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.,* Keen, *et al.,* 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing

25   gradient gel electrophoresis (DGGE). *See, e.g.,* Myers, *et al.,* 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control

30   and sample DNA. *See, e.g.,* Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the

known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g.,* Saiki, *et al.,* 1986. *Nature* 324: 163; Saiki, *et al.,* 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of

5   different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of

10 interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g.,* Gibbs, *et al.,* 1989. *Nucl. Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.,* Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the

15 mutation to create cleavage-based detection. *See, e.g.,* Gasparini, *et al.,* 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g.,* Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known

20 mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.,* in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving A

25 NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

30

PHARMACOGENOMICS

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (*e.g.,* NOVX gene expression), as identified by a screening assay described herein

can be administered to individuals to treat (prophylactically or therapeutically) disorders (The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various

5   dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.) In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can

10   lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic

15   regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

     Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons.

20   See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These

25   pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

30   As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome Pregnancy Zone Protein Precursor enzymes CYP2D6 and CYP2C19)

has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence

5   of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no

10   therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

        Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or

15   mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can

20   avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with A NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

MONITORING OF EFFECTS DURING CLINICAL TRIALS

25   Monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of NOVX (e.g., the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity,

30   can be monitored in clinical trails of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical

trails of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune

5   responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) that modulates NOVX activity (e.g., identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for

10   example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of

15   activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the

20   effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of A NOVX protein, mRNA, or genomic DNA in the

25   preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration

30   sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be

desirable to decrease expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

5   The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis,

10   subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, AIDS, bronchial asthma, Crohn's disease;

15   multiple sclerosis, treatment of Albright Hereditary Ostoeodystrophy, and other diseases, disorders and conditions of the like.

         These methods of treatment will be discussed more fully, below.

DISEASES AND DISORDERS

20   Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (*i*) an aforementioned peptide, or analogs,

25   derivatives, fragments or homologs thereof; (*ii*) antibodies to an aforementioned peptide; (*iii*) nucleic acids encoding an aforementioned peptide; (*iv*) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous

30   recombination (*see, e.g.,* Capecchi, 1989. *Science* 244: 1288-1292); or (*v*) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that

5   may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or

10   mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

15

PROPHYLACTIC METHODS

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to the subject an agent that modulates NOVX expression or at least one

20   NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the NOVX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its

25   progression. Depending upon the type of NOVX aberrancy, for example, A NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

30   ### Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of

NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of A NOVX protein, a peptide, A NOVX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or

5   more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by

10   culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of A NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination

15   of agents that modulates (*e.g.*, up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering A NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situ*ations in which NOVX is

20   abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (*e.g.*, preclampsia).

25

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

30   In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice,

chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

5   **Prophylactic and Therapeutic Uses of the Compositions of the Invention**

The NOVX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders including, but not limited to: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's

10   Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.

As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in

15   need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias.

20   Both the novel nucleic acid encoding the NOVX protein, and the NOVX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.,* some peptides have been found to possess anti-bacterial properties). These materials are further useful in the generation of

25   antibodies, which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

30

EXAMPLES

Example A: Polynucleotide And Polypeptide Sequences, And Homology Data

EXAMPLE 1.

5   The NOV1 clone was analyzed, and the nucleotide and encoded polypeptide
sequences are shown in Table 1A.

Table 1A. NOV1 Sequence Analysis		
	SEQ ID NO: 1	1794 bp
NOV1a, CG100041-01 DNA Sequence	AGAAGCTCCCGAGTGTCCGGCCTAGAGGCCATGAGAAGGCAGTGGGGGTCTGCCATGA GGGCGGCCGAGCAGGCGGGCTGCATGGTGAGCGCCTCCCGGGCCGGACAGCCCGAGGC GGGCCCGTGGAGCTGCAGCGGGGTAATCCTGAGCCGTAGCCCGGGCCTGGTGCTTTGC CACGGGGGCATCTTCGTCCCCTTCCTGCGAGCTGGCAGCGAAGTCCTGACCCGGCCCG GCGCCGTCTTCCTGCCTGGCGACAGTTGCAGGGACGACCTGCGCCTGCACGTGCAGTG GGCCCCAACGGCAGCTGCCCCGTCTGGGAAGTGGGGAGTGCCTCTGCCCGGCCGCCCC GTCTGGGAAGTGAGGAGTGCCTCTGCCCGGCCGCCCATCATCTGGGATGTGAGGAGCG CCCCTCCTTCTCCTCCTCCTCCTCCCCTCCTCCTCCTGCTTGTTTCCCTTTGCTCCTT CTTTTTAGCCCACTTTAGTCTAAAATATGGATCATGTTTTAAAAATACATTTTATTTT GTTAGAGCAGCCCAGGCAGACACTAAGACGCTCACAGAACAGGAGGGAAGTTTAAAGA CGCTGGGCTGGTTTGCGCTGCTGGGCGTGCGGCTAGGCCAGGAAGAGTGGAGGAGACG CGGGCCAACGATGGCGGTGTCGCCTCTCGGGGCCGTGCCCAAGGGTGCGCCATTGCTG GTCTGCGGCTCCCCTTTCGGCGCCTTCTGCCCCGACATCTTTCTCAACACGCTGAGCT GCGGGGTGCTCAGCAACGTGGCCGGCCCACTGCTGCTTACCGACGCACGCTGCCTGCC CGGCACCGAGGGCGGCGGCGTGTTCACCGCGCGGCCCGCGGGGGCGCTGGTGGCGCTG GTGGTGGCGCCGCTCTGTTGGAAGGCCGGCGAATGGGTGGGCTTCACGCTGCTCTGCG CCGCCGCCCCCCTTTTCCGCGCCGCCCGCGACGCGCTTCACCGCCTGCCGCACAGCAC CGCTGCCCTGGCCGCCCTTCTGCCGCCAGAGGTGGGCGTCCCGTGGGGTCTGCCCCTC CGAGACTCCGGGCCCCTGTGGGCAGCCGCGGCAGTGTTGGTGGAGTGCGGCACCGTAT GGGGCTCCGGAGTGGCTGTGGCACCCCGCCTTGTAGTGACCTGTCGGCACGTGTCCCC TCGGAAGCAGCCAGGGTCCTGGTGCGCTCCACCACCCCCAAGAGTGTGGCCATCTGG GGCCGTGTGGTATTTGCCACTCAGGAGACATGTCCCTATGACATAGCAGTGGTGAGCC TGGAGGAGGACCTGGATGATGTCCCCATCCCTGTGCCCGCTGAGCACTTCCATGAAGG CGAGGCTGTGAGTGTGGTGGGCTTTGGCGTCTTTGGCCAGTCTTGCGGGCCCTCGGTG ACCTCAGGCATCCTTTCGGCTGTGGTGCAGGTGAATGGCACGCCCGTAATGCTGCAGA CCACGTGTGCTGTGCACAGCGGCTCCAGTGGGGGACCCCTCTTCTCCAACCACTCAGG AAACCTCCTTGGTATAATCACCAGCAACACCCGGGACAATAATACGGGGGCCACCTAC CCCCACCTGAACTTCAGCATTCCCATCACGGTGCTCCAGCCGGCCCTGCAGCAGTACA GCCAGACCCAAGACCTAGGTGGCCTCCGTGAGCTGGACCGCGCTGCTGAGCCAGTCAG GGTGGTGTGGCGGTTGCAGCGGCCCCTGGCAGAGGCCCCGCGGAGCAAGCTCTGAGGC TGTGTTACCACCTTTGGAAAGAAGAGTGACCTTTTTCTGCTGTAGGAAGTGATG
	ORF Start: ATG at 31	ORF Stop: TGA at 1735
	SEQ ID NO: 2	568 aa	MW at 60004.6kD
NOV1a, CG100041-01 Protein Sequence	MRRQWGSAMRAAEQAGCMVSASRAGQPEAGPWSCSGVILSRSPGLVLCHGGIFVPFLR AGSEVLTRPGAVFLPGDSCRDDLRLHVQWAPTAAAPSGKWGVPLPGRPVWEVRSASAR PPIIWDVRSAPPSPPPPPLLLLLLVSLCSFFLAHFSLKYGSCFKNTFYFVRAAQADTKT LTEQEGSLKTLGWFALLGVRLGQEEWRRRGPTMAVSPLGAVPKGAPLLVCGSPFGAFC PDIFLNTLSCGVLSNVAGPLLLTDARCLPGTEGGGVFTARPAGALVALVVAPLCWKAG EWVGFTLLCAAAPLFRAARDALHRLPHSTAALAALLPPEVGVPWGLPLRDSGPLWAAA AVLVECGTVWGSGVAVAPRLVVTCRHVSPREAARVLVRSTTPKSVAIWGRVVFATQET CPYDIAVVSLEEDLDDVPIPVPAEHFHEGEAVSVVGFGVFGQSCGPSVTSGILSAVVQ VNGTPVMLQTTCAVHSGSSGGPLFSNHSGNLLGIITSNTRDNNTGATYPHLNFSIPIT VLQPALQQYSQTQDLGGLRELDRAAEPVRVVWRLQRPLAEAPRSKL	

Further analysis of the NOV1a protein yielded the following properties shown in
Table 1B.

Table 1B. Protein Sequence Properties NOV1a	
PSort analysis:	0.8741 probability located in microbody (peroxisome); 0.8266 probability located in mitochondrial inner membrane; 0.6500 probability located in plasma membrane; 0.3000 probability located in Golgi body
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV1a protein against the Geneseq database, a proprietary
database that contains sequences published in patents and patent publication, yielded

5   several homologous proteins shown in Table 1C.

Table 1C. Geneseq Results for NOV1a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB56819	Human prostate cancer antigen protein sequence SEQ ID NO:1397 - Homo sapiens, 204 aa. [WO200055174-A1, 21-SEP-2000]	365..568 1..204	204/204 (100%) 204/204 (100%)	e-115
AAM95577	Human reproductive system related antigen SEQ ID NO: 4235 - Homo sapiens, 136 aa. [WO200155320-A2, 02-AUG-2001]	365..500 1..136	124/136 (91%) 124/136 (91%)	6e-65
AAO08663	Human polypeptide SEQ ID NO 22555 - Homo sapiens, 91 aa. [WO200164835-A2, 07-SEP-2001]	101..125 43..68	23/26 (88%) 24/26 (91%)	1e-05
AAG98947	E. coli growth and proliferation related protein sequence SEQ ID NO:417 - Escherichia coli, 355 aa. [WO200134810-A2, 17-MAY-2001]	359..520 80..237	51/170 (30%) 80/170 (47%)	4e-05
AAY75748	Neisseria gonorrheae ORF 986 protein sequence SEQ ID NO:2968 - Neisseria gonorrheae, 499 aa. [WO9957280-A2, 11-NOV-1999]	337..534 92..293	62/218 (28%) 104/218 (47%)	9e-05

In a BLAST search of public sequence databases, the NOV1a protein was found to
have homology to the proteins shown in the BLASTP data in Table 1D.

Table 1D. Public BLASTP Results for NOV1a				
Protein Accession Number	Protein/Organism/Length	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9DBA6	1300019N10RIK PROTEIN - Mus musculus (Mouse), 568 aa.	1..568 1..568	419/569 (73%) 463/569 (80%)	0.0
Q96AR5	SIMILAR TO RIKEN CDNA 1300019N10 GENE - Homo sapiens (Human), 435 aa (fragment).	137..568 3..435	395/433 (91%) 403/433 (92%)	0.0
Q8VZD4	AT1G28320/F3H9_2 - Arabidopsis thaliana (Mouse-ear cress), 709 aa.	435..568 517..663	57/149 (38%) 79/149 (52%)	2e-18
Q9FZA5	F3H9.3 PROTEIN - Arabidopsis thaliana (Mouse-ear cress), 688 aa.	435..568 496..642	57/149 (38%) 79/149 (52%)	2e-18
Q95RU7	LD11031P - Drosophila melanogaster (Fruit fly), 509 aa.	221..559 184..503	99/353 (28%) 157/353 (44%)	3e-16

PFam analysis predicts that the NOV1a protein contains the domains shown in the Table 1E.

Table 1E. Domain Analysis of NOV1a			
Pfam Domain	NOV1a Match Region	Identities/ Similarities for the Matched Region	Expect Value
trypsin	408..528	26/157 (17%) 85/157 (54%)	0.0044

<u>EXAMPLE 2.</u>

The NOV2 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 2A.

Table 2A. NOV2 Sequence Analysis		
	SEQ ID NO: 3	2274 bp
NOV2a, CG105716-01 DNA Sequence	ATGGTCCCCGACACCGCCTGCGTTCTTCTGCTCACCCTGGCTGCCCTCGGCGCGTCCG GACAGGGCCAGAGCCCGTTGGGCTCAGACCTGGGCCCGCAGATGCTTCGGGAACTGCA GGAAACCAACGCGGCGCTGCAGGACGTGCGGGACTGGCTGCGGCAGCAGGTCAGGGAG ATCACGTTCCTGAAAAACACGGTGATGGAGTGTGACGCGTGCGGGATGCAGCAGTCAG TACGCACCGGCCTACCCAGCGTGCGGCCCCTGCTCCACTGCGCGCCCGGCTTCTGCTT CCCCGGCGTGGCCTGCATCCAGACGGAGAGCGGCGGCCGCTGCGGCCCCTGCCCCGCG GGCTTCACGGGCAACGGCTCGCACTGCACCGACGTCAACGAGTGCAACGCCCACCCCT GCTTCCCCCGAGTCCGCTGTATCAACACCAGCCCGGGGTTCCGCTGCGAGGCTTGCCC GCCGGGGTACAGCGGCCCCACCCACCAGGGCGTGGGGCTGGCTTTCGCCAAGGCCAAC AAGCAGGTTTGCACGGACATCAACGAGTGTGAGACCGGGCAACATAACTGCGTCCCCA ACTCCGTGTGCATCAACACCCGGGGCTCCTTCCAGTGCGGCCCCGTGCCAGCCCGGCTT CGTGGGCGACCAGGCGTCCGGCTGCCAGCGGCGCGCACAGCGCTTCTGCCCCGACGGC TCGCCCAGCGAGTGCCACGAGCATGCAGACTGCGTCCTAGAGCGCGATGGCTCGCGGT CGTGCGTGTGTGCCGTTGGCTGGGCCGGCAACGGGATCCTCTGTGGTCGCGACACTGA CCTAGACGGCTTCCCGGACGAGAAGCTGCGCTGCCCGGAGCGCCAGTGCCGTAAGGAC AACTGCGTGACTGTGCCCAACTCAGGGCAGGAGGATGTGGACCGCGATGGCATCGGAG ACGCCTGCGATCCGGATGCCGACGGGGACGGGGTCCCCAATGAAAAGGACAACTGCCC GCTGGTGCGGAACCCAGACCAGCGCAACACGGACGAGGACAAGTGGGGCGATGCGTGC GACAACTGCCGGTCCCAGAAGAACGACGACCAAAAGGACACAGACCAGGACGGCCGGG GCGATGCGTGCGACGACGACATCGACGGCGACCGGATCCGCAACCAGGCCGACAACTG CCCTAGGGTACCCAACTCAGACCAGAAGGACAGTGATGGCGATGGTATAGGGGATGCC TGTGACAACTGTCCCCAGAAGAGCAACCCGGATCAGGCGGATGTGGACCACGACTTTG TGGGAGATGCTTGTGACAGCGATCAAGACCAGGATGGAGACGGACATCAGGACTCTCG GGACAACTGTCCCACGGTGCCTAACAGTGCCCAGGAGGACTCAGACCACGATGGCCAG GGTGATGCCTGCGACGACGACGACGACAATGACGGAGTCCCTGACAGTCGGGACAACT GCCGCCTGGTGCCTAACCCCGGCCAGGAGGACGCGGACAGGGACGGCGTGGGCGACGT GTGCCAGGACGACTTTGATGCAGACAAGGTGGTAGACAAGATCGACGTGTGTCCGGAG AACGCTGAAGTCACGCTCACCGACTTCAGGGCCTTCCAGACAGTCGTGCTGGATCCTG AAGGGGATGCCCAGATCGATCCCAACTGGGTGGTCCTGAACCAGGGCATGGAGATTGT ACAGACCATGAACAGTGATCCTGGCCTGGCAGTGGGGTACACAGCTTTTAATGGAGTT GACTTCGAAGGGACCTTCCATGTGAATACCCAGACAGATGATGACTATGCAGGCTTTA TCTTTGGCTACCAAGATAGCTCCAGCTTCTACGTGGTCATGTGGAAGCAGACGGAGCA GACATATTGGCAAGCCACCCCATTCCGAGCAGTTGCAGAACCTGGCATTCAGCTCAAG GCTGTGAAGTCTAAGACAGGTCCAGGGGAGCATCTCCGGAACGCTCTGTGGCATACAG GAGACACAGAGTCCCAGGTGCGGCTGCTGTGGAAGGACCCGCGAAACGTGGGTTGGAA GGACAAGAAGTCCTATCGTTGGTTCCTGCAGCACCGGCCCCAAGTGGGCTACATCAGG GTGCGATTCTATGAGGGCCCTGAGCTGGTGGCCGACAGCAACGTGGTCTTGGACACAA CCATGCGGGGTGGCCGCCTGGGGGTCTTCTGCTTCTCCCAGGAGAACATCATCTGGGC CAACCTGCGTTACCGCTGCAATGACACCATCCCAGAGGACTATGAGACCCATCAGCTG CGGCAAGCCTAG
	ORF Start: ATG at 1	ORF Stop: TAG at 2272
	SEQ ID NO: 4	757 aa ‖ MW at 82915.7kD
NOV2a, CG105716-01 Protein Sequence	MVPDTACVLLLTLAALGASGQGQSPLGSDLGPQMLRELQETNAALQDVRDWLRQQVRE ITFLKNTVMECDACGMQQSVRTGLPSVRPLLHCAPGFCFPGVACIQTESGGRCGPCPA GFTGNGSHCTDVNECNAHPCFPRVRCINTSPGFRCEACPPGYSGPTHQGVGLAFAKAN KQVCTDINECETGQHNCVPNSVCINTRGSFQCGPCQPGFVGDQASGCQRRAQRFCPDG SPSECHEHADCVLERDGSRSCVCAVGWAGNGILCGRDTDLDGFPDEKLRCPERQCRKD NCVTVPNSGQEDVDRDGIGDACDPDADGDGVPNEKDNCPLVRNPDQRNTDEDKWGDAC

| | DNCRSQKNDDQKDTDQDGRGDACDDDIDGDRIRNQADNCPRVPNSDQKDSDGDGIGDA
CDNCPQKSNPDQADVDHDFVGDACDSDQDQDGDGHQDSRDNCPTVPNSAQEDSDHDGQ
GDACDDDDDNDGVPDSRDNCRLVPNPGQEDADRDGVGDVCQDDFDADKVVDKIDVCPE
NAEVTLTDFRAFQTVVLDPEGDAQIDPNWVVLNQGMEIVQTMNSDPGLAVGYTAFNGV
DFEGTFHVNTQTDDDYAGFIFGYQDSSSFYVVMWKQTEQTYWQATPFRAVAEPGIQLK
AVKSKTGPGEHLRNALWHTGDTESQVRLLWKDPRNVGWKDKKSYRWFLQHRPQVGYIR
VRFYEGPELVADSNVVLDTTMRGGRLGVFCFSQENIIWANLRYRCNDTIPEDYETHQL
RQA |

Further analysis of the NOV2a protein yielded the following properties shown in Table 2B.

Table 2B. Protein Sequence Properties NOV2a	
PSort analysis:	0.5278 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 21 and 22

A search of the NOV2a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded

5   several homologous proteins shown in Table 2C.

Table 2C. Geneseq Results for NOV2a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB00044	Human cartilage oligomeric matrix protein (COMP) - Homo sapiens, 757 aa. [WO200044908-A2, 03-AUG-2000]	1..757 1..757	747/757 (98%) 747/757 (98%)	0.0
AAR56248	Xenopus thrombospondin-4 - Xenopus laevis, 889 aa. [WO9413794-A, 23-JUN-1994]	29..756 150..886	522/740 (70%) 605/740 (81%)	0.0
AAR56249	Human thrombospondin-4 - Homo sapiens, 961 aa. [WO9413794-A, 23-JUN-1994]	16..753 211..952	524/752 (69%) 603/752 (79%)	0.0
AAM93335	Human polypeptide, SEQ ID NO: 2869 - Homo sapiens, 762 aa. [EP1130094-A2, 05-SEP-2001]	26..750 26..751	485/738 (65%) 575/738 (77%)	0.0
AAM79078	Human protein SEQ ID NO 1740 - Homo sapiens, 776 aa. [WO200157190-A2, 09-AUG-2001]	16..577 211..776	363/576 (63%) 436/576 (75%)	0.0

In a BLAST search of public sequence databases, the NOV2a protein was found to have homology to the proteins shown in the BLASTP data in Table 2D.

Table 2D. Public BLASTP Results for NOV2a				
Protein Accession Number	Protein/Organism/Length	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P49747	Cartilage oligomeric matrix protein precursor (COMP) - Homo sapiens (Human), 757 aa.	1..757 1..757	748/757 (98%) 748/757 (98%)	0.0
O14592	COMP_HUMAN - Homo sapiens (Human), 817 aa.	1..742 1..742	733/742 (98%) 734/742 (98%)	0.0
Q9BG80	CARTILAGE OLIGOMERIC MATRIX PROTEIN - Equus caballus (Horse), 755 aa.	1..757 1..755	692/757 (91%) 711/757 (93%)	0.0
P35444	Cartilage oligomeric matrix protein precursor (COMP) - Rattus norvegicus (Rat), 755 aa.	5..757 4..755	680/753 (90%) 706/753 (93%)	0.0
Q9R0G6	CARTILAGE OLIGOMERIC MATRIX PROTEIN PRECURSOR - Mus musculus (Mouse), 755 aa.	5..756 4..754	678/752 (90%) 705/752 (93%)	0.0

PFam analysis predicts that the NOV2a protein contains the domains shown in the Table 2E.

Table 2E. Domain Analysis of NOV2a			
Pfam Domain	**NOV2a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
EGF	229..266	10/47 (21%) 30/47 (64%)	0.007
tsp_3	300..314	11/15 (73%) 13/15 (87%)	0.02
tsp_3	336..350	9/15 (60%) 15/15 (100%)	0.057
tsp_3	359..373	10/15 (67%) 15/15 (100%)	0.22
tsp_3	395..409	12/15 (80%) 15/15 (100%)	0.014
tsp_3	418..432	10/15 (67%) 13/15 (87%)	0.042
tsp_3	456..470	12/15 (80%) 15/15 (100%)	0.25
tsp_3	492..506	10/15 (67%) 14/15 (93%)	0.2

EXAMPLE 3.

The NOV3 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 3A.

Table 3A. NOV3 Sequence Analysis		
	SEQ ID NO: 5	799 bp
NOV3a, CG113569-01 DNA Sequence	CGGCCGCGTCGACCGGGCCCCGAGGCACAGCCAGGGCACCAGGTGGAGCACCAGCTAC GCGTGGCGCAGCGCAGCGTCCCTAGCACCGAGCCTCCCGCAGCCGCCGAGATGCTGCG AACAGAGAGCTGCCGCCCCAGGTCGCCCGCCGGACAGGTGGCCGCGGCGTCCCCGCTC CTGCTGCTGCTGCTGCTGCTCGCCTGGTGCGCGGGCGCCTGCCGAGGTGCTCCAATAT TACCTCAAGGATTACAGCCTGAACAACAGCTACAGTTGTGGAATGAGGCATCCAACGC ACTGGAGGAGCTTTGCTTTATGATTATGGGAATGCTACCAAAGCCTCAGGAACAAGAT GAAAAAGATAATACTAAAAGGTTCTTATTTCATTATTCGAAGACACAGAAGTTGGGCA AGTCAAATGTTGTGTCGTCAGTTGTGCATCCGTTGCTGCAGCTCGTTCCTCACCTGCA TGAGAGAAGAATGAAGAGATTCAGAGTGGACGAAGAATTCCAAAGTCCCTTTGCAAGT CAAAGTCGAGGATATTTTTTATTCAGGCCACGGAATGGAAGAAGGTCAGCAGGGTTCA TTTAAAATGGATGCCAGCTAATTTTCCACAGAGCAATGCTATGGAATACAAAATGTAC TGACATTTTGTTTTCTTCTGAAAAAAAATCCTTGCTAAATGTACTCTGTTGAAAATCC CTGTGTTGTCAATGTTCTCAGTTGTAACAATGTTGTAAATGTTCAATTTGTTGAAAAT TAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACCAAAAAAAT
	ORF Start: ATG at 109	ORF Stop: TAA at 583
	SEQ ID NO: 6	158 aa	MW at 18002.8kD
NOV3a, CG113569-01 Protein Sequence	MLRTESCRPRSPAGQVAAASPLLLLLLLLLAWCAGACRGAPILPQGLQPEQQLQLWNEA SNALEELCFMIMGMLPKPQEQDEKDNTKRFLFHYSKTQKLGKSNVVSSVVHPLLQLVP HLHERRMKRFRVDEEFQSPFASQSRGYFLFRPRNGRRSAGFI	
	SEQ ID NO: 7	746 bp
NOV3b, CG113569-03 DNA Sequence	AGTCCTGCGTCCGGGCCCCGAGGCACAGCCAGGGCACCAGGTGGAGCACCAGCTACGC GTGGCGCAGCGCAGCGTCCCTAGCACCGAGCCTCCCGCAGCCGCCGAGATGCTGCGAA CAGAGAGCTGCCGCCCCAGGTCGCCCGCCGGACAGGTGGCCGCGGCGTCCCCGCTCCT GCTGCTGCTGCTGCTGCTCGCCTGGTGCGCGGGCGCCTGCCGAGGTGCTCCAATATTA CCTCAAGGATTACAGCCTGAACAACAGCTACAGTTGTGGAATGAGGCATCCAACGCAC TGGAGGAGCTTTGCTTTATGATTATGGGAATGCTACCAAAGCCTCAGGAACAAGATGA AAAAGATAATACTAAAAGGTTCTTATTTCATTATTCGAAGACACAGAAGTTGGGCAAG TCAAATGTTGTGTCGTCAGTTGTGCATCCGTTGCTGCAGCTCGTTCCTCACCTGCATG AGAGAAGAATGAAGAGATTCAGAGTGGACGAAGAATTCCAAAGTCCCTTTGCAAGTCA AAGTCGAGGATATTTTTTATTCAGGCCACGGAATGGAAGAAGGTCAGCAGGGTTCATT TAAAATGGATGCCAGCTAATTTTCCACAGAGCAATGCTATGGAATACAAAATGTACTG ACATTTTGTTTTCTTCTGAAAAAAAATCCTTGCTAAATGTACTCTGTTGAAAATCCCTG TGTTGTCAATGTTCTCAGTTGTAACAATGTTGTAAATGTTCAATTTGTTG
	ORF Start: ATG at 107	ORF Stop: TAA at 581
	SEQ ID NO: 8	158 aa	MW at 18002.8kD
NOV3b, CG113569-03 Protein Sequence	MLRTESCRPRSPAGQVAAASPLLLLLLLLLAWCAGACRGAPILPQGLQPEQQLQLWNEA SNALEELCFMIMGMLPKPQEQDEKDNTKRFLFHYSKTQKLGKSNVVSSVVHPLLQLVP HLHERRMKRFRVDEEFQSPFASQSRGYFLFRPRNGRRSAGFI	

Sequence comparison of the above protein sequences yields the following

5   sequence relationships shown in Table 3B.

Table 3B. Comparison of NOV3a against NOV3b.		
Protein Sequence	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV3b	1..158 1..158	118/158 (74%) 118/158 (74%)

Further analysis of the NOV3a protein yielded the following properties shown in Table 3C.

Table 3C. Protein Sequence Properties NOV3a	
PSort analysis:	0.8200 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Cleavage site between residues 39 and 40

A search of the NOV3a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 3D.

Table 3D. Geneseq Results for NOV3a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE14264	Human NMU-25 peptide - Homo sapiens, 25 aa. [WO200181418-A1, 01-NOV-2001]	126..150 1..25	25/25 (100%) 25/25 (100%)	1e-07
AAE03630	Human neuromedin U neuropeptide (NMU)-25 bioactive peptide - Homo sapiens, 25 aa. [WO200144297-A1, 21-JUN-2001]	126..150 1..25	25/25 (100%) 25/25 (100%)	1e-07
AAB99193	Human neuromedin U-25 peptide - Homo sapiens, 25 aa. [WO200140797-A1, 07-JUN-2001]	126..150 1..25	25/25 (100%) 25/25 (100%) ·	1e-07
AAG63360	Amino acid sequence of a human neuromedin U peptide - Homo sapiens, 25 aa. [WO200157524-A1, 09-AUG-2001]	126..150 1..25	25/25 (100%) 25/25 (100%)	1e-07
AAB91380	Tachykinins peptide SEQ ID NO:556 - Homo sapiens, 25 aa. [WO200069900-A2, 23-NOV-2000]	126..150 1..25	19/25 (76%) 21/25 (84%)	2e-04

5

In a BLAST search of public sequence databases, the NOV3a protein was found to have homology to the proteins shown in the BLASTP data in Table 3E.

Table 3E. Public BLASTP Results for NOV3a				
Protein Accession Number	**Protein/Organism/Length**	**NOV3a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
P48645	Neuromedin U-25 precursor (NmU-25) - Homo sapiens (Human), 174 aa.	1..158 1..174	158/174 (90%) 158/174 (90%)	2e-85
Q9QXK8	Neuromedin U-23 precursor (NmU-23) - Mus musculus (Mouse), 174 aa.	1..158 1..174	117/176 (66%) 127/176 (71%)	4e-55
P12760	Neuromedin U-23 precursor (NmU-23) - Rattus norvegicus (Rat), 174 aa.	1..158 1..174	113/176 (64%) 126/176 (71%)	3e-51
P34965	Neuromedin U-25 (NmU-25) - Oryctolagus cuniculus (Rabbit), 25 aa.	126..150 1..25	22/25 (88%) 23/25 (92%)	2e-05
Q9JLQ5	NEUROMEDIN U - Mus musculus (Mouse), 29 aa (fragment).	131..158 2..29	20/28 (71%) 23/28 (81%)	8e-05

PFam analysis predicts that the NOV3a protein contains the domains shown in the Table 3F.

Table 3F. Domain Analysis of NOV3a			
Pfam Domain	**NOV3a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
NMU	126..150	21/25 (84%) 25/25 (100%)	2.4e-17

5

<u>EXAMPLE 4.</u>

The NOV4 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 4A.

Table 4A. NOV4 Sequence Analysis		
	SEQ ID NO: 9	1397 bp
NOV4a, CG56602-02 DNA Sequence	TCGCTGCCGCCGCTCCCAGCAACAAATCACAGCTGCATTTTCTCGTTGATCCAGTCGA TGTAGGCGGACACCCGGGTGTAGACTACCGGCTTCTTGCGGGTGTTGCAGCCCCGCCG GGAGCCAGTCCTGAGCACCTAAC**ATG**TTGGGCATCACTGTCCTCGCTGCGCTCTCCG CCAGTGCCTCCAGCTGTGGGGTGCCCAGCTTCCCGCCCAACCTATCCGCCCGAGTGGT GGGAGGAGAGGATGCCCGGCCCCACAGCTGGCCCTGGCAGGTAAGCCTGCTCCAGTAC CTCAAGAACGACACGTGGAGGCATACGTGTGGTGGGACTTTGATTGCTAGCAACTTCG TCCTCACTGCCGCCCACTGCATCAGCAACACCCGGACCTACCGTGTGGCCGTGGGAAA GAACAACCTGGAGGTGGAAGACGAAGAAGGATCCCTGTTTGTGGGTGTGGACACCATC CACGTCCACAAGAGATGGAATGCCTTGGATTCCAGCAATGATATTGCCCTCATCAAGC TTGCAGAGCATGTGGAGCTGAGTGACACCATCCAGGTGGCCTGCCTGCCAGAGAAGGA CTCCCTGCTCCCCAAGGACTACCCCTGCTATGTCACCGGCTGGGGCCGCCTCAACGGC CCCATTGCTGATAAGCTGCAGCAGGGCCTGCAGCCCGTGGTGGATCACGCCACGTGCT CCAGGATTGACTGGTGGGGCTTCAGGGTGAAGAAAACCATGGTGTGCGCTGGGGGCGA TGGCGTCATCTCAGCCTGCAATGGGGACTCCGGTGGCCCACTGAACTGCCAGTTGGAG AACGGTTCCTGGGAGGTGTTTGGCATCGTCAGCTTTGGCTCCCGGCGGGGCTGCAACA CCCGCAAGAAGCCGGTAGTCTACACCCGGGTGTCCGCCTACATCGACTGGATCAACGA GGTGGGTGCTGCCTCCACAGCTGTCCCTGCACCTGTCAGCCCCTCCCCCTCACTCACC CATCCCCTCACTCATTCACTCATTCATGCGTTTATTCATTCATTCATTTATTCACTCA TTCATGCATTTATTCACTCATTCATGCATTCATTCATTTATTCACTTATTCAGTCACT CATTCATGTATTTATTCATTTATTCATTCACTCATGCATTTATTCATTCATTCATTTA TTCACTTATTCATTCACTCATTCATGTATTCATTCATTCATTCATGCATTTATTTACT CATTCATCCATTTATTCACTCATTCATTTGCTCATTCAGTGATTCATTCATGCACTTC TTCACACATTCACTCTCTCATTCAAGTAATATTGGTTGAGTGCTTCCAG**TAG**CAGGCC TTGAGTTGGGTGCCAATAAGGAAACAGTCATTGACTCCTCCATCCATCCATTCACTGC CTTCA
	ORF Start: ATG at 141	ORF Stop: TAG at 1326
	SEQ ID NO: 10	395 aa	MW at 43653.5kD
NOV4a, CG56602-02 Protein Sequence	MLGITVLAALSASASSCGVPSFPPNLSARVVGGEDARPHSWPWQVSLLQYLKNDTWRH TCGGTLIASNFVLTAAHCISNTRTYRVAVGKNNLEVEDEEGSLFVGVDTIHVHKRWNA LDSSNDIALIKLAEHVELSDTIQVACLPEKDSLLPKDYPCYVTGWGRLNGPIADKLQQ GLQPVVDHATCSRIDWWGFRVKKTMVCAGGDGVISACNGDSGGPLNCQLENGSWEVFG IVSFGSRRGCNTRKKPVVYTRVSAYIDWINEVGAASTAVPAPVSPSPSLTHPLTHSLI HAFIHSFIYSLIHAFIHSFMHSFIYSLIQSLIHVFIHLFIHSCIYSFIHLFTYSFTHS CIHSFIHAFIYSFIHLFTHSFAHSVIHSCTSSHIHSLIQVILVECFQ

5

Further analysis of the NOV4a protein yielded the following properties shown in Table 4B.

Table 4B. Protein Sequence Properties NOV4a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.2409 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 17 and 18

A search of the NOV4a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 4C.

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAR90683	Human caldecrin contg. preprosequence - Rattus sp, 268 aa. [WO9600287-A, 04-JAN-1996]	1..263 1..264	255/265 (96%) 256/265 (96%)	e-150
AAR88481	Human elastase IV protein - Homo sapiens, 268 aa. [WO9601270-A1, 18-JAN-1996]	1..263 1..264	253/265 (95%) 254/265 (95%)	e-148
AAY51839	Human elastase IV homolog HEIV protein fragment - Homo sapiens, 268 aa. [US6030791-A, 29-FEB-2000]	1..263 1..264	252/265 (95%) 253/265 (95%)	e-148
AAW89410	Human homologue of rat elastase IV - Homo sapiens, 268 aa. [US5856109-A, 05-JAN-1999]	1..263 1..264	252/265 (95%) 253/265 (95%)	e-148
AAW40530	Human elastase homologue HEIV protein - Homo sapiens, 268 aa. [US5738991-A, 14-APR-1998]	1..263 1..264	252/265 (95%) 253/265 (95%)	e-148

In a BLAST search of public sequence databases, the NOV4a protein was found to have homology to the proteins shown in the BLASTP data in Table 4D.

Table 4D. Public BLASTP Results for NOV4a

Protein Accession Number	Protein/Organism/Length	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q99895	Caldecrin precursor (EC 3.4.21.2) (Chymotrypsin C) - Homo sapiens (Human), 268 aa.	1..263 1..264	256/265 (96%) 257/265 (96%)	e-151
S68826	pancreatic elastase (EC 3.4.21.36) isoform 2 precursor - human, 268 aa.	1..263 1..264	255/265 (96%) 256/265 (96%)	e-150
P55091	Caldecrin precursor (EC 3.4.21.2) (Chymotrypsin C) (Serum calcium-decreasing factor) - Rattus norvegicus (Rat), 268 aa.	1..263 1..264	202/265 (76%) 229/265 (86%)	e-119
JQ1473	pancreatic elastase (EC 3.4.21.36) IV precursor - rat, 268 aa.	1..263 1..264	188/265 (70%) 212/265 (79%)	e-106
P08217	Elastase 2A precursor (EC 3.4.21.71) - Homo sapiens (Human), 269 aa.	1..264 1..265	169/268 (63%) 203/268 (75%)	4e-93

PFam analysis predicts that the NOV4a protein contains the domains shown in the Table 4E.

Table 4E. Domain Analysis of NOV4a

Pfam Domain	NOV4a Match Region	Identities/ Similarities for the Matched Region	Expect Value
trypsin	30..261	110/264 (42%) 190/264 (72%)	1.4e-88

EXAMPLE 5.

The NOV5 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 5A.

Table 5A. NOV5 Sequence Analysis		
	SEQ ID NO: 11	3210 bp
NOV5a, CG57415-01 DNA Sequence	CAATTCTATTCGCTTGTTATTGGACTTGAAACTCCCTTTGACCTCGGAAACTGAAGAT GAGGTTGCCATGGGAACTGCTGGTACTGCAATCATTCATTTTGTGCCTTGCAGATGAT TCCACACTGCATGGCCCGATTTTTATTCAAGAACCAAGTCCTGTAATGTTCCCTTTGG ATTCTGAGGAGAAAAAAGTGAAGCTCAATTGTGAAGTTAAAGGAAATCCAAAACCTCA TATCAGGTGGAAGTTAAATGGAACAGATGTTGACACTGGTATGGATTTCCGCTACAGT GTTGTTGAAGGGAGCTTGTTGATCAATAACCCCAATAAAACCCAAGATGCTGGAACGT ACCAGTGCACAGCGACAAACTCGTTTGGAACAATTGTTAGCAGAGAAGCAAAGCTTCA GTTTGCTTATCTTGACAACTTTAAAACAAGAACAAGAAGCACTGTGTCTGTCCGTCGA GGTCAAGGAATGGTGCTACTGTGTGGCCCGCCACCCCATTCTGGAGAGCTGAGTTATG CCTGGATCTTCAATGAATACCCTTCCTATCAGGATAATCGCCGCTTTGTTTCTCAAGA GACTGGGAATCTGTATATTGCCAAAGTAGAAAAATCAGATGTTGGGAATTATACCTGT GTGGTTACCAATACCGTGACAAACCACAAGGTCCTGGGGCCACCTACACCACTAATAT TGAGAAATGATGGAGTGATGGGTGAATATGAGCCCAAAATAGAAGTGCAGTTCCCAGA AACAGTTCCGACTGCAAAAGGAGCAACGGTGAAGCTGGAATGCTTTGCTTTAGGAAAT CCAGTACCAACTATTATCTGGCGAAGAGCTGATGGAAAGCCAATAGCAAGGAAAGCCA GAAGACACAAGTCAAATGGAATTCTTGAGATCCCTAATTTTCAGCAGGAGGATGCTGG TTTATATGAATGTGTAGCTGAAAATTCCAGAGGGAAAAATGTAGCAAGGGGACAGCTA ACTTTCTATGCTCAACCTAATTGGATTCAAAAAATAAATGATATTCACGTGGCCATGG AAGAAAATGTCTTTTGGGAATGTAAAGCAAATGGAAGGCCTAAGCCTACATACAAGTG GCTAAAAAATGGCGAACCTCTGCTAACTCGGGATAGAATTCAAATTGAGCAAGGAACA CTCAACATAACAATAGTGAACCTCTCAGATGCTGGCATGTATCAGTGTTTGGCAGAGA ATAAACATGGAGTTATCTTTTTCCAACGCAGAGCTTAGTGTTATAGCTGTAGGTCCAGA TTTTTCAAGAACACTCTTGAAAAGAGTAACTCTTGTCAAAGTGGGAGGTGAAGTTGTC ATTGAGTGTAAGCCAAAAGCGTCTCCAAAACCTGTTTACACCTGGAAGAAAGGAAGGG ATATATTAAAAGAAAATGAAAGAATTACCATTTCTGAAGATGGAAACCTCAGAATCAT CAACGTTACTAAATCAGACGCTGGGAGTTATACCTGTATAGCCACTAACCATTTTGGA ACTGCTAGCAGTACTGGAACTTGGTAGTGAAAGATCCAACAAGGGTAATGGTACCCC CTTCCAGTATGGATGTCACTGTTGGAGAGAGTATTGTTTTACCGTGCCAGGTAACGCA TGATCACTCGCTAGACATCGTGTTTACTTGGTCATTTAATGGACACCTGATAGACTTT GACAGAGATGGGGACCACTTTGAAAGAGTTGGAGGGCAGGATTCAGCTGGTGATTTGA TGATCCGAAACATCCAACTGAAGCATGCTGGGAAATATGTCTGCATGGTCCAAACAAG TGTGGACAGGCTATCTGCTGCTGCAGACCTGATTGTAAGAGGTCCTCCAGGTCCCCCA GAGGCTGTGACAATAGACGAAATCACAGATACCACTGCTCAGCTCTCCTGGAGACCCG GGCCTGACAACCACAGCCCCATCACCATGTATGTCATTCAAGCCAGGACTCCATTCTC CGTGGGCTGGCAAGCAGTCAGTACAGTCCCAGAACTCATTGATGGGAAGACATTCACA GCGACCGTGGTGGGTTTGAACCCTTGGGTTGAATATGAATTCCGCACAGTTGCAGCCA ACGTGATTGGGATTGGGGAGCCCAGCCGCCCCTCAGAGAAACGGAGAACAGAAGAAGC TCTCCCCGAAGTCACACCAGCGAATGTCAGTGGTGGCGGAGGCAGCAAATCTGAACTG GTTATAACCTGGGAGACGGTCCCTGAGGAATTACAGAATGGTCGAGGCTTTGGTTATG TGGTGGCCTTCCGGCCCTACGGTAAAATGATCTGGATGCTGACAGTGCTGGCCTCAGC TGATGCCTCTAGATACGTGTTCAGGAATGAGAGCGTGCACCCCTTCTCTCCCTTTGAG GTTAAAGTAGGTGTCTTCAACAACAAAGGAGAAGGCCCTTTCAGTCCCACCACGGTGG TGTATTCTGCAGAAGAAGAACCCACCAAACCACCAGCCAGTATCTTTGCCAGAAGTCT TTCTGCCACAGATATTGAAGTTTTCTGGGCCTCCCCACTGGAGAAGAATAGAGGACGA ATACAAGGTTATGAGGTTAAATATTGGAGACATGAAGACAAAGAAGAAAATGCTAGAA AAATACGAACAGTTGGAAATCAGACATCAACAAAAATCACGAACTTAAAAGGCAGTGT GCTGTATCACTTAGCTGTCAAGGCATATAATTCTGCTGGGACAGGCCCCTCTAGTGCA ACAGTCAATGTGACAACCCGAAAGCCACCACCAAGTCAACCCCCGGGAACATCATAT GGAATTCATCAGACTCCAAAATTATTCTGAATTGGGATCAAGTGAAGGCCCTGGATAA TGAGTCGGAAGTAAAAGGATACAAAGTAGTCTTGTACAGATGGAACAGACAAAGCAGC

	ACATCTGTCATTGAAACAAATAAAACATCGGTGGAGCTTTCTTTGCCTTTCGATGAAG ATTATATAATAGAAATTAAGCCATTCAGCGACGGAGGAGATGGCAGCAGCAGTGAACA AATTCGAATTCCAAAGATATCAAATGCCTACGCGAGAGGATCTGGGGCTTCCACTTCG AATGCATGTACGCTGTCAGCCATCAGTACAATAATGATTTCCCTCACAGCTAGGTCCA GTTTATGACAAAAGTTATCTGAAGGACTTGCTGTTTATAATATAAGCAACATTTAGCT AGTTGTTTTGAAGACACCCA
	ORF Start: ATG at 57	ORF Stop: TGA at 3138
	SEQ ID NO: 12	1027 aa	MW at 113551.7kD
NOV5a, CG57415-01 Protein Sequence	MRLPWELLVLQSFILCLADDSTLHGPIFIQEPSPVMFPLDSEEKKVKLNCEVKGNPKP HIRWKLNGTDVDTGMDFRYSVVEGSLLINNPNKTQDAGTYQCTATNSFGTIVSREAKL QFAYLDNFKTRTRSTVSVRRGQGMVLLCGPPPHSGELSYAWIFNEYPSYQDNRRFVSQ ETGNLYIAKVEKSDVGNYTCVVTNTVTNHKVLGPPTPLILRNDGVMGEYEPKIEVQFP ETVPTAKGATVKLECFALGNPVPTIIWRRADGKPIARKARRHKSNGILEIPNFQQEDA GLYECVAENSRGKNVARGQLTFYAQPNWIQKINDIHVAMEENVFWECKANGRPKPTYK WLKNGEPLLTRDRIQIEQGTLNITIVNLSDAGMYQCLAENKHGVIFSNAELSVIAVGP DFSRTLLKRVTLVKVGGEVVIECKPKASPKPVYTWKKGRDILKENERITISEDGNLRI INVTKSDAGSYTCIATNHFGTASSTGNLVVKDPTRVMVPPSSMDVTVGESIVLPCQVT HDHSLDIVFTWSFNGHLIDFDRDGDHFERVGGQDSAGDLMIRNIQLKHAGKYVCMVQT SVDRLSAAADLIVRGPPGPPEAVTIDEITDTTAQLSWRPGPDNHSPITMYVIQARTPF SVGWQAVSTVPELIDGKTFTATVVGLNPWVEYEFRTVAANVIGIGEPSRPSEKRRTEE ALPEVTPANVSGGGGSKSELVITWETVPEELQNGRGFGYVVAFRPYGKMIWMLTVLAS ADASRYVFRNESVHPFSPFEVKVGVFNNKGEGPFSPTTVVYSAEEEPTKPPASIFARS LSATDIEVFWASPLEKNRGRIQGYEVKYWRHEDKEENARKIRTVGNQTSTKITNLKGS VLYHLAVKAYNSAGTGPSSATVNVTTRKPPPSQPPGNIIWNSSDSKIILNWDQVKALD NESEVKGYKVVLYRWNRQSSTSVIETNKTSVELSLPFDEDYIIEIKPFSDGGDGSSSE QIRIPKISNAYARGSGASTSNACTLSAISTIMISLTARSSL

Further analysis of the NOV5a protein yielded the following properties shown in
Table 5B.

Table 5B. Protein Sequence Properties NOV5a	
PSort analysis:	0.3700 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1800 probability located in nucleus; 0.1134 probability located in microbody (peroxisome)
SignalP analysis:	Cleavage site between residues 19 and 20

A search of the NOV5a protein against the Geneseq database, a proprietary
database that contains sequences published in patents and patent publication, yielded
5   several homologous proteins shown in Table 5C.

Table 5C. Geneseq Results for NOV5a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB53276	Human polypeptide #16 - Homo sapiens, 1026 aa. [WO200181363-A1, 01-NOV-2001]	13..1027 14..1026	1009/1015 (99%) 1011/1015 (99%)	0.0

AAW29667	Homo sapiens DL185_1 clone secreted protein - Homo sapiens, 1028 aa. [WO9830695-A2, 16-JUL-1998]	1..1000 1..1001	617/1003 (61%) 775/1003 (76%)	0.0
AAU18339	Human endocrine polypeptide SEQ ID No 294 - Homo sapiens, 447 aa. [WO200155364-A2, 02-AUG-2001]	582..1027 3..447	445/446 (99%) 445/446 (99%)	0.0
AAM43534	Human polypeptide SEQ ID NO 212 - Homo sapiens, 456 aa. [WO200155308-A2, 02-AUG-2001]	578..1027 8..456	442/450 (98%) 444/450 (98%)	0.0
AAR87028	Human contactin - Homo sapiens, 1018 aa. [WO9535373-A2, 28-DEC-1995]	25..986 40..989	439/965 (45%) 603/965 (61%)	0.0

In a BLAST search of public sequence databases, the NOV5a protein was found to have homology to the proteins shown in the BLASTP data in Table 5D.

Table 5D. Public BLASTP Results for NOV5a				
Protein Accession Number	Protein/Organism/Length	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q62845	NEURAL CELL ADHESION PROTEIN BIG-2 PRECURSOR - Rattus norvegicus (Rat), 1026 aa.	1..1027 1..1026	970/1027 (94%) 1012/1027 (98%)	0.0
Q07409	NEURONAL GLYCOPROTEIN - Mus musculus (Mouse), 1028 aa.	1..1019 1..1020	665/1022 (65%) 833/1022 (81%)	0.0
Q62682	BIG-1 PROTEIN PRECURSOR - Rattus norvegicus (Rat), 1028 aa.	1..1019 1..1020	666/1022 (65%) 829/1022 (80%)	0.0
AAH26119	SIMILAR TO AXONAL-ASSOCIATED CELL ADHESION MOLECULE - Homo sapiens (Human), 697 aa.	329..1027 1..697	697/699 (99%) 697/699 (99%)	0.0
CAA67504	BRAIN-DERIVED IMMUNOGLOBULIN SUPERFAMILY MOLECULE - Mus musculus (Mouse), 705 aa.	1..697 1..697	645/697 (92%) 675/697 (96%)	0.0

PFam analysis predicts that the NOV5a protein contains the domains shown in the Table 5E.

Table 5E. Domain Analysis of NOV5a			
Pfam Domain	**NOV5a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Ig	43..102	17/62 (27%) 44/62 (71%)	8.7e-08
Ig	137..196	15/62 (24%) 39/62 (63%)	0.00015
Ig	240..297	17/61 (28%) 44/61 (72%)	9.1e-10
Ig	337..386	15/53 (28%) 37/53 (70%)	0.00015
Ig	422..479	16/61 (26%) 42/61 (69%)	1.3e-09
Ig	512..578	16/68 (24%) 44/68 (65%)	1.1e-06
fn3	597..686	30/91 (33%) 69/91 (76%)	5.4e-18
fn3	699..789	22/92 (24%) 59/92 (64%)	0.074
fn3	801..889	27/90 (30%) 62/90 (69%)	1.8e-11
fn3	901..985	19/89 (21%) 59/89 (66%)	0.17

EXAMPLE 6.

The NOV6 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 6A.

Table 6A. NOV6 Sequence Analysis		
	SEQ ID NO: 13	5115 bp
NOV6a, CG58504-01 DNA Sequence	GAATTCCGGGAGCGGGCGGGCTGCGAGGCCGCGGGGCATGCGGGAGGCGGAGGGGTGG GACCGGGTGGCTGCGCCCATTCCACACCCGCCGAAAGCGGACACTGTCAGCTGAATCA CTCCCCTTTTAGGAGGAGGGAGGGGGAAAAGGTGTCTAGCTAATTTCTGCTTAAAAAA GCACAGGAGATCGCGGGTCAGCTTTGCAGTCGCTGCCTTCTCGCGCCTGACCATGCAC CCCTGCATCTTCCTGCTGGGCACAGGCGAGCGCTTTATTTCTGGAGCTGAGGGCTAAA ACTTTTTTTCACTTTTTCTTCTCCTCAACATCTGAATCATGCCATGTGCCCAGAGGAGCT GGCTTGCAAACCTTTCCGTGGTGGCTCAGCTCCTTAACTTTGGGGCGCTTTGCTATGG GAGACAGCCTCAGCCAGGCCCGGTTCGCTTCCCGGACAGGAGGCAAGAGCATTTTATC AAGGGCCTGCCAGAATACCACGTGGTGGGTCCAGTCCGAGTAGATGCCAGTGGGCATT TTTTGTCATATGGCTTGCACTATCCCATCACGAGCAGCAGGAGGAAGAGAGATTTGGA TGGCTCAGAGGACTGGGTGTACTACAGAATTTCTCACGAGGAGAAGGACCTGTTTTTT AACTTGACGGTCAATCAAGGATTTCTTTCCAATAGCTACATCATGGAGAAGAGATATG GGAACCTCTCCCATGTTAAGATGATGGCTTCCTCTGCCCCCCCTCTGCCATCTCAGTGG CACGGTTCTACAGCAGGGCACCAGAGTTGGGACGGCCAGCCCTCAGTGCCTGCCATGGA CTGACTGGATTTTTCCAACTACCACATGGAGACTTTTTCATTGAACCCGTGAAGAAGC ATCCACTGGTTGAGGGAGGGTACCACCCGCACATCGTTTACAGGAGGCAGAAAGTTCC AGAAACCAAGGAGCCAACCTGTGGATTAAAGGACAGTGTTAACATCTCCCAGAAGCAA GAGCTATGGCGGGAGAAGTGGGAGAGGCACAACTTGCCAAGCAGAAGCCTCTCTCGGC GTTCCATCAGCAAGGAGAGATGGGTGGAGACACTGGTGGTGGCCGACACAAAGATGAT TGAATACCATGGGAGTGAGAATGTGGAGTCCTACATCCTCACCATCATGAACATGGTC ACTGGGTTGTTCCATAACCCAAGCATTGGCAATGCAATTCACATTGTTGTGGTTCGGC TCATTCTACTCGAAGAAGAAGAGCAAGGACTGAAAATAGTTCACCATGCAGAAAAGAC ACTGTCTAGCTTCTGCAAGTGGCAGAAGAGTATCAATCCCAAGAGTGACCTCAATCCT GTTCATCACGACGTGGCTGTCCTTCTCACCAGAAAGGACATCTGTGCTGGTTTCAATC GCCCCTGCGAGACCCTGGGCCTGTCTCACCTTTCAGGAATGTGTCAGCCTCACCGCAG TTGTAACATCAATGAAGATTCGGGACTCCCTCTGGCTTTCACAATTGCCCATGAGCTA GGACACAGCTTCGGCATCCAGCATGATGGGAAAGAAAATGACTGTGAGCCTGTGGGCA GACATCCGTACATCATGTCCCGCCAGCTCCAGTACGATCCCACTCCGCTGACATGGTC CAAGTGCAGCGAGGAGTACATCACCCGCTTCTTGGACCGAGGCTGGGGGTTCTGTCTT GATGACATACCTAAAAAGAAAGGCTTGAAGTCCAAGGTCATTGCCCCCGGAGTGATCT ATGATGTTCACCACCAGTGCCAGCTACAATATGGACCCAATGCTACCTTCTGCCAGGA AGTAGAAAACGTCTGCCAGCACTGTGGTGCTCCGTGAAGGGCTTTTGTCGCTCTAAG CTGGACGCTGCTGCAGATGGAACTCAATGTGGTGAGAAGAAGTGGTGTATGGCAGGCA AGTGCATCACAGTGGGGAAGAAACCAGAGAGCATTCCTGGAGGCTGGGGCCGCTGGTC ACCCTGGTCCCACTGTTCCAGGACCTGTGGGGCTGGAGTCCAGAGCGCAGAGAGGCTC TGCAACAACCCCGAGCCAAAGTTTGGAGGGAAATATTGCACTGGAGAAAGAAAACGCT ATCGCTTGTGCAACGTCCACCCCTGTCGCTCAGAGGCACCAACATTTCGGCAGATGCA GTGCAGTGAATTTGACACTGTTCCCTACAAGAATGAACTCTACCACTGGTTTCCCATT TTTAACCCAGCACATCCTTGTGAGCTCTACTGCCGACCCATAGATGGCCAGTTTTCTG AGAAAATGCTGGATGCTGTCATTGATGGTACCCCTTGCTTTGAAGGCGGCAACAGCAG AAATGTCTGTATTAATGGCATATGTAAGATGGTTGGCTGTGACTATGAGATCGATTCC AATGCCACCGAGGATCGCTGCGGTGTGTGCCTGGGAGATGGCTCTTCCTGCCAGACTG TGAGAAAGATGTTTAAGCAGAAGGAAGGATCTGGTTATGTTGACATTGGGCTCATTCC AAAAGGAGCAAGGGACATAAGAGTGATGGAAATTGAGGGAGCTGGAAACTTCCTGGCC ATCAGGAGTGAAGATCCTGAAAAATATTACCTGAATGGAGGGTTTATTATCCAGTGGA ACGGGAACTATAAGCTGGCAGGGACTGTCTTTCAGTATGACAGGAAAGGAGACCTGGA AAAGCTGATGGCCACAGGTCCCACCAATGAGTCTGTGTGGATCCAGCTTCTATTCCAG GTGACTAACCCTGGCATCAAGTATGAGTACACAATCCAGAAAGATGGCCTTGACAATG ATGTTGAGCAGATGTACTTCTGGCAGTACGGCCACTGGACAGAGTGCAGTGTGACCTG CGGGACAGGTATCCGCCGCCAAACTGCCCATTGCATAAAGAAGGGCCGCGGGATGGTG

110

	AAAGCTACATTCTGTGACCCAGAAACACAGCCCAATGGGAGACAGAAGAAGTGCCATG
	AAAAGGCTTGTCCACCCAGGTGGTGGGCAGGGGAGTGGGAAGCATGCTCGGCGACATG
	CGGGCCCCACGGGGAGAAGAAGCGAACCGTGCTGTGCATCCAGACCATGGTCTCTGAC
	GAGCAGGCTCTCCCGCCCACAGACTGCCAGCACCTGCTGAAGCCCAAGACCCTCCTTT
	CCTGCAACAGAGACATCCTGTGCCCCTCGGACTGGACAGTGGGCAACTGGAGTGAGTG
	TTCTGTTTCCTGTGGTGGTGGAGTGCGGATTCGCAGTGTCACATGTGCCAAGAACCAT
	GATGAACCTTGCGATGTGACAAGGAAACCCAACAGCCGAGCTCTGTGTGGCCTCCAGC
	AATGCCCTTCTAGCCGGAGAGTTCTGAAACCAAACAAAGGCACTATTTCCAATGGAAA
	AAACCCACCAACACTAAAGCCCGTCCCTCCACCTACATCCAGGCCCAGAATGCTGACC
	ACACCCACAGGGCCTGAGTCTATGAGCACAAGCACTCCAGCAATCAGCAGCCCTAGTC
	CTACCACAGCCTCCAAAGAAGGAGACCTGGGTGGGAAACAGTGGCAAGATAGCTCAAC
	CCAACCTGAGCTGAGCTCTCGCTATCTCATTTCCACTGGAAGCACTTCCCAGCCCATC
	CTCACTTCCCAATCCTTGAGCATTCAGCCAAGTGAGGAAATGTTTCCAGTTCAGATA
	CTGGTCCTACCTCGGAGGGAGGCCTTGTAGCTACAACAACAAGTGGTTCTGGCTTGTC
	ATCTTCCCGCAACCCTATCACTTGGCCTGTGACTCCATTTTACAATACCTTGACCAAA
	GGTCCAGAAATGGAGATTCACAGTGGCTCAGGGGAAGAAAGAGAACAGCCTGAGGACA
	AAGATGAAAGCAATCCTGTAATATGGACCAAGATCAGAGTACCTGGAAATGACGCTCC
	AGTGGAAAGTACAGAAATGCCACTTGCACCTCCACTAACACCAGATCTCAGCAGGGAG
	TCCTGGTGGCCACCCTTCAGCACAGTAATGGAAGGACTGCTCCCCAGCCAAAGGCCCA
	CTACTTCCGAAACTGGGACACCCAGAGTTGAGGGGATGGTTACTGAAAAGCCAGCCAA
	CACTCTGCTCCCTCTGGGAGGAGACCACCAGCCAGAACCCTCAGGAAAGACGGCAAAC
	CGTAACCACCTGAAACTTCCAAACAACATGAACCAAACAAAAAGTTCTGAACCAGTCC
	TGACTGAGGAGGATGCAACAAGTCTGATTACTGAGGGCTTTTTGCTAAATGCCTCCAA
	TTACAAGCAGCTCACAAACGGCCACGGCTCTGCACACTGGATCGTCGGAAACTGGAGC
	GAGTGCTCCACCACATGTGGCCTGGGGGCCTACTGGAAAAGGGTGGAGTGCACCACCC
	AGATGGATTCTGACTGTGCGGCCATCCAGAGACCTGACCCTGCAAAAAGATGCCACCT
	CCGTCCCTGTGCTGGCTGGAAAGTGGGAAACTGGAGCAAGTGCTCCAGAAACTGCAGT
	GGGGGCTTCAAGATACGCGAGATTCAGTGCGTGGACAGCCGGGACCACCGGAACCTGA
	GGCCATTTCACTGCCAGTTCCTGGCCGGCATTCCTCCCCCATTGAGCATGAGCTGTAA
	CCCGGAGCCCTGTGAGGCGTGGCAGGTGGAGCCTTGGAGCCAGTGCTCCAGGTCCTGT
	GGAGGTGGAGTTCAGGAGAGAGGAGTGTTCTGTCCAGGAGGCCTCTGTGATTGGACAA
	AAAGACCCACATCCACCATGTCTTGCAATGAGCACCTGTGCTGTCACTGGGCCACTGG
	GAACTGGGACCTGTGTTCCACTTCCTGTGGAGGTGGCTTTCAGAAGAGGATTGTCCAA
	TGTGTGCCCTCAGAGGGCAATAAAACTGAAGACCAAGACCAATGTCTATGTGATCACA
	AACCCAGACCTCCAGAATTCAAAAAATGCAACCAGCAGGCCTGCAAGAAAAGTGCCGA
	TTTACTTTGCACTAAGGACAAACTGTCAGCCAGTTTCTGCCAGACACTGAAAGCCATG
	AAGAAATGTTCTGTGCCCACCGTGAGGGCTGAGTGCTGCTTCTCGTGTCCCCAGACAC
	ACATCACACACACCCAAAGGCAAAGAAGGCAACGGTTGCTCCAAAAGTCAAAAGAACT
	CTAAGCCCAAA

	ORF Start: ATG at 327		ORF Stop: TAA at 5106
	SEQ ID NO: 14	1593 aa	MW at 177543.9kD
NOV6a, CG58504-01 Protein Sequence	MPCAQRSWLANLSVVVAQLLNFGALCYGRQPQPGPVRFPDRRQEHFIKGLPEYHVVGPV RVDASGHFLSYGLHYPITSSRRKRDLDGSEDWVYYRISHEEKDLFFNLTVNQGFLSNS YIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGTAALSACHGLTGFFQLPHGDF FIEPVKKHPLVEGGYHPHIVYRRQKVPETKEPTCGLKDSVNISQKQELWREKWERHNL PSRSLSRRSISKERWVETLVVADTKMIEYHGSENVESYILTIMNMVTGLFHNPSIGNA IHIVVVRLILLEEEEQGLKIVHHAEKTLSSFCKWQKSINPKSDLNPVHHDVAVLLTRK DICAGFNRPCETLGLSHLSGMCQPHRSCNINEDSGLPLAFTIAHELGHSFGIQHDGKE NDCEPVGRHPYIMSRQLQYDPTPLTWSKCSEEYITRFLDRGWGFCLDDIPKKKGLKSK VIAPGVIYDVHHQCQLQYGPNATFCQEVENVCQTLWCSVKGFCRSKLDAAADGTQCGE KKWCMAGKCITVGKKPESIPGGWGRWSPWSHCSRTCGAGVQSAERLCNNPEPKFGGKY CTGERKRYRLCNVHPCRSEAPTFRQMQCSEFDTVPYKNELYHWFPIFNPAHPCELYCR PIDGQFSEKMLDAVIDGTPCFEGGNSRNVCINGICKMVGCDYEIDSNATEDRCGVCLG DGSSCQTVRKMFKQKEGSGYVDIGLIPKGARDIRVMEIEGAGNFLAIRSEDPEKYYLN GGFIIQWNGNYKLAGTVFQYDRKGDLEKLMATGPTNESVWIQLLFQVTNPGIKYEYTI QKDGLDNDVEQMYFWQYGHWTECSVTCGTGIRRQTAHCIKKGRGMVKATFCDPETQPN GRQKKCHEKACPPRWWAGEWEACSATCGPHGEKKRTVLCIQTMVSDEQALPPTDCQHL LKPKTLLSCNRDILCPSDWTVGNWSECSVSCGGGVRIRSVTCAKNHDEPCDVTRKPNS

	RALCGLQQCPSSRRVLKPNKGTISNGKNPPTLKPVPPPTSRPRMLTTPTGPESMSTST PAISSPSPTTASKEGDLGGKQWQDSSTQPELSSRYLISTGSTSQPILTSQSLSIQPSE ENVSSSDTGPTSEGGLVATTTSGSGLSSSRNPITWPVTPFYNTLTKGPEMEIHSGSGE EREQPEDKDESNPVIWTKIRVPGNDAPVESTEMPLAPPLTPDLSRESWWPPFSTVMEG LLPSQRPTTSETGTPRVEGMVTEKPANTLLPLGGDHQPEPSGKTANRNHLKLPNNMNQ TKSSEPVLTEEDATSLITEGFLLNASNYKQLTNGHGSAHWIVGNWSECSTTCGLGAYW KRVECTTQMDSDCAAIQRPDPAKRCHLRPCAGWKVGNWSKCSRNCSGGFKIREIQCVD SRDHRNLRPFHCQFLAGIPPPLSMSCNPEPCEAWQVEPWSQCSRSCGGGVQERGVFCP GGLCDWTKRPTSTMSCNEHLCCHWATGNWDLCSTSCGGGFQKRIVQCVPSEGNKTEDQ DQCLCDHKPRPPEFKKCNQQACKKSADLLCTKDKLSASFCQTLKAMKKCSVPTVRAEC CFSCPQTHITHTQRQRRQRLLQKSKEL	
	SEQ ID NO: 15	252 bp
NOV6b, 169648407 DNA Sequence	AAGCTTCACCAGTGCCAGCTACAATATGGACCCAATGCTACCTTCTGCCAGGAAGTAG AAAACGTCTGCCAGACACTGTGGTGCTCCGTGAAGGGCTTTTGTCGCTCTAAGCTGGA CGCTGCTGCAGATGGAACTCAATGTGGTGAGAAGAAGTGGTGTATGGCAGGCAAGTGC ATCACAGTGGGGAAGAAACCAGAGAGCATTCCTGGAGGCTGGGGCCGCTGGTCACCCT GGTCCCACTGTTCCCTCGAG	
	ORF Start: at 1	ORF Stop: end of sequence
	SEQ ID NO: 16	84 aa	MW at 9323.7kD
NOV6b, 169648407 Protein Sequence	KLHQCQLQYGPNATFCQEVENVCQTLWCSVKGFCRSKLDAAADGTQCGEKKWCMAGKC ITVGKKPESIPGGWGRWSPWSHCSLE	
	SEQ ID NO: 17	252 bp
NOV6c, 169648441 DNA Sequence	AAGCTTCACCAGTGCCAGCTACAATATGGACCCAATGCTACCTTCTGCCAGGAAGTAG AAAACGTCTGCCAGACACTGTGGTGCTCCGTGAAGGGCTTTTGTCGCTCTAAGCTGGA CGCTGCTGCAGATGGAACTCAATGTGGTGAGAAGAAGTGGTGTATGGCAGGCAAGTGC ATCACAGTGGGGAAGAAACCAGAGAGCATTCCTGGAGGCTGCGGCCGCTGGTCACCCT GGTCCCACTGTTCCCTCGAG	
	ORF Start: at 1	ORF Stop: end of sequence
	SEQ ID NO: 18	84 aa	MW at 9240.6kD
NOV6c, 169648441 Protein Sequence	KLHQCQLQYGPNATFCQEVENVCQTLWCSVKGFCRSKLDAAADGTQCGEKKWCMAGKC ITVGKKPESIPGGCGRWSPWSHCSLE	

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 6B.

Table 6B. Comparison of NOV6a against NOV6b and NOV6c.		
Protein Sequence	**NOV6a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV6b	476..555 3..82	80/80 (100%) 80/80 (100%)
NOV6c	476..555 3..82	79/80 (98%) 79/80 (98%)

Further analysis of the NOV6a protein yielded the following properties shown in Table 6C.

Table 6C. Protein Sequence Properties NOV6a	
PSort analysis:	0.5087 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 26 and 27

5

A search of the NOV6a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 6D.

Table 6D. Geneseq Results for NOV6a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV6a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB74944	Human ADAM type metal protease MDTS1 protein SEQ ID NO:1 - Homo sapiens, 1686 aa. [JP2001008687-A, 16-JAN-2001]	24..1574 21..1673	733/1718 (42%) 941/1718 (54%)	0.0
AAE00913	Human 27875 ADAM-TS protein, alternative version - Homo sapiens, 1686 aa. [WO200131034-A1, 03-MAY-2001]	24..1574 21..1673	731/1718 (42%) 939/1718 (54%)	0.0
AAE00934	Human 27875 ADAM-TS (a disintegrin and metalloproteinase) - Homo sapiens, 1686 aa. [WO200131034-A1, 03-MAY-2001]	24..1574 21..1673	731/1718 (42%) 939/1718 (54%)	0.0
AAB86949	Human metalloprotease MPTS-19 protein - Homo sapiens, 1690 aa. [DE10107360-A1, 06-SEP-2001]	24..1574 25..1677	731/1718 (42%) 938/1718 (54%)	0.0
AAB72283	Human ADAMTS-7 amino acid sequence - Homo sapiens, 997 aa. [WO200111074-A2, 15-FEB-2001]	24..903 21..936	483/935 (51%) 609/935 (64%)	0.0

5

In a BLAST search of public sequence databases, the NOV6a protein was found to have homology to the proteins shown in the BLASTP data in Table 6E.

Table 6E. Public BLASTP Results for NOV6a				
Protein Accession Number	Protein/Organism/Length	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P58397	ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM- TS12) - Homo sapiens (Human), 1593 aa.	1..1593 1..1593	1593/1593 (100%) 1593/1593 (100%)	0.0
CAC38921	SEQUENCE 2 FROM PATENT WO0131034 - Homo sapiens (Human), 1686 aa.	24..1574 21..1673	731/1718 (42%) 939/1718 (54%)	0.0
Q9UKP4	ADAMTS-7 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 7) (ADAM-TS 7) (ADAM-TS7) - Homo sapiens (Human), 997 aa.	24..903 21..936	485/935 (51%) 611/935 (64%)	0.0
CAD20434	SEQUENCE 8 FROM PATENT WO0188156 - Homo sapiens (Human), 1044 aa (fragment).	40..1001 27..1008	373/1011 (36%) 540/1011 (52%)	0.0
Q9H324	ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) - Homo sapiens (Human), 1077 aa (fragment).	40..1001 1..982	372/1011 (36%) 540/1011 (52%)	0.0

PFam analysis predicts that the NOV6a protein contains the domains shown in the Table 6F.

Table 6F. Domain Analysis of NOV6a			
Pfam Domain	**NOV6a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Pep_M12B_propep	105..222	27/128 (21%) 82/128 (64%)	0.00058
Reprolysin	246..456	65/224 (29%) 149/224 (67%)	1.3e-15
tsp_1	546..596	23/54 (43%) 38/54 (70%)	3.1e-13
tsp_1	827..881	15/65 (23%) 39/65 (60%)	0.041
tsp_1	945..995	17/59 (29%) 39/59 (66%)	9.2e-05
tsp_1	1314..1364	13/57 (23%) 29/57 (51%)	0.027
tsp_1	1426..1471	14/55 (25%) 32/55 (58%)	0.044
tsp_1	1474..1530	14/64 (22%) 40/64 (62%)	0.0029

EXAMPLE 7.

The NOV7 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 7A.

Table 7A. NOV7 Sequence Analysis		
	SEQ ID NO: 19	4324 bp
NOV7a, CG58586-01 DNA Sequence	GCCCAATAAATGTGAACATGGGATCTGTCACGGGAGCTGTCCTCAAGACGCTACTTCT GTTATCTACTCAAAATTGGAACAGAGTCGAAGCTGGGAATTCCTATGACTGTGATGAT CCTCTTGTGTCTGCCTTGCCTCAGGCATCCTTCAGCAGTTCTTCCGAGCTCTCCAGCA GTCATGGTCCTGGATTTGCAAGGCTGAATAGAAGAGATGGAGCTGGTGGCTGGTCTCC ACTTGTGTCTAACAAATACCAGTGGTTGCAGATTGACCTTGGAGAGAGAATGGAGGTC ACCGCTGTGGCCACTCAAGGGGGATATGGTAGCTCCAACTGGGTGACCAGCTACCTCC TGATGTTCAGTGATAGTGGCTGGAACTGGAAACAATATCGCCAAGAGGACAGCATCTG GGGTTTTTCAGGAAATGCAAATGCAGACAGTGTTGTGTACTATAGACTCCAGCCTTCT ATCAAAGCCAGATTTCTGCGCTTCATCCCTTTGGAATGGAACCCCAAGGGCAGAATTG GAATGCGAATCGAAGTGTTCGGATGTGCATACAGATCAGAAGTGGTTGATCTTGATGG AAAAAGTTCCCTTCTCTACAGATTTGATCAAAAATCCCTGAGCCCAATAAAAGACATT ATTTCTTTGAAATTCAAAACCATGCAGAGTGATGGGATTCTACTCCACAGGGAAGGGC CAAATGGAGATCACATCACACTGCAATTAAGAAGAGCAAGACTCTTTTTACTTATTAA TTCAGGTGAAGCTAAACTGCCTTCCACTTCCACCCTGGTCAATCTCACCCTGGGCAGC CTGCTAGATGATCAGCATTGGCATTCAGTGCTCATCCAGCGTTTGGGCAAACAAGTCA ACTTCACAGTGGACGAACACAGGCATCATTTCCATGCACGGGGAGAATTCAATCTCAT GAATCTTGATTATGAGATCAGCTTTGGAGGGATTCCAGCACCTGGAAAATCAGTGTCA TTCCCACATAGAAATTTTCATGGATGTTTAGAAAATCTCTATTATAATGGAGTGGATA TCATTGATTTGGCCAAGCAGCAAAAACCACAGATCATTGCTATGGGAAATGTGTCATT TTCTTGTTCACAACCACAATCTATGCCCGTGACTTTTCTGAGCTCCAGGAGTTATTTA GCACTGCCAGACTTCTCTGGAGAGGAGGAGGTTTCTGCCACTTTTCAATTTCGAACTT GGAATAAGGCAGGGCTTCTGCTGTTCAGTGAACTTCAGCTGATTTCAGGGGGTATCCT CCTCTTTCTGAGTGATGGAAAACTTAAGTCGAATCTCTACCAGCCAGGAAAATTACCC AGTGACATCACAGCAGGTGTCGAATTAAATGATGGGCAGTGGCATTCTGTCTCTTTAT CTGCTAAAAAGAATCACTTGAGTGTGGCGGTGGACGGCCAGATGGCTTCTGCTGCTCC TCTGCTGGGGCCTGAGCAGATTTATTCGGGTGGCACCTATTATTTTGGAGGTTGTCCT GACAAAAGCTTTGGATCCAAATGTAAAAGTCCACTTGGTGGATTTCAGGGATGTATGA GGCTCATTTCTATCAGCGGCAAAGTGGTAGATCTGATTTCAGTTCAGCAGGGGTCCCT TGGGAACTTCAGTGACCTTCAGATAGACTCATGTGGCATCTCAGACAGGTGTTTGCCC AACTATTGTGAACACGGTGGGGAGTGTTCCCAGTCCTGGAGCACCTTTCATTGTAACT GTACCAACACTGGTTACAGAGGAGCTACTTGCCATAACTCTATCTATGAGCAGTCATG TGAAGCCTATAAGCACAGAGGAAATACTTCAGGGTTTTACTATATAGATTCAGATGGA AGTGGTCCCCTGGAACCATTTCTTCTATATTGCAATATGACCGAAACTGCATGGACCA TCATACAGCACAACGGCTCTGACTTAACAAGAGTCAGAAATACTAATCCAGAGAACCC ATATGCTGGGTTTTTCGAGTATGTGGCCAGCATGGAGCAACTTCAGGCCACTATTAAC CGTGCAGAGCACTGTGAACAGGAGTTTACTTATTACTGCAAGAAGTCACGGCTGGTCA ATAAGCAAGATGGAACCCCTCTGAGTTGGTGGGTAGGAAGAACCAATGAAACGCAAAC CTACTGGGGAGGTTCTTCGCCTGATCTTCAAAAATGTACTTGTGGATTAGAGGGAAAC TGCATTGATTCTCAGTATTACTGCAATTGTGATGCTGACCGGAATGAATGGACCAATG ACACTGGATTGCTTGCTTATAAAGAACATCTTCCAGTAACTAAGATCGTGATTACAGA CACAGGCCGACTGCATTCAGAAGCAGCTTATAAACTGGGGCCTCTGCTCTGCCGGGGA GACAGATCATTTTGGAATTCAGCTTCCTTTGATACCGAGGCTTCATATCTTCATTTTC CTACCTTCCACGGAGAACTTAGCGCGGATGTATCTTTCTTTTTTAAGACAACAGCTTC ATCTGGGGTATTTTTAGAGAACTTGGGGATTGCTGATTTTATACGGATAGAGCTTCGC TCTCCGACAGTAGTGACTTTTTCATTTGATGTGGGGAATGGGCCTTTTGAAATCTCAG TGCAGTCACCCACCCACTTCAACGACAACCAGTGGCACCATGTGAGGGTTGAAAGGAA CATGAAGGAGGCCTCCCTTCAAGTGGATCAGCTGACACCAAAGACACAGCCCGCCCCC GCTGATGGGCATGTCCTGTTACAGCTCAACAGTCAGCTCTTCGTGGGTGGAACGGCCA CCAGACAGAGAGGCTTTCTGGGCTGCATTCGGTCTCTGCAGTTGAATGGGATGACCCT GGATTTGGAAGAAAGAGCCCAGGTGACTCCAGAAGTGCAGCCAGGTTGTAGGGGACAT

117

TGCAGCAGCTATGGGAAGTTATGCCGCAATGGAGGGAAATGCAGAGAAAGACCCATTG
GGTTCTTTTGTGACTGCACTTTCTCTGCATACACAGGGCCATTCTGCTCAAATGAGAT
TTCTGCATATTTTGGATCTGGCTCATCCGTGATATACAATTTTCAAGAAAATTATCTT
TTAAGTAAAAACTCCAGCTCCACGCTGCTTCATTTCATGGTGATATGAAGCTGAGCA
GAGAAATGATCAAATTTAGTTTCCGAACAACACGAACACCAAGCTTGCTGCTTTTTGT
GAGCTCCTTTTACAAAGAATACCTTTCTGTGATCATTGCCAAAAATGGAAGTTTGCAG
ATCAGGTACAAGTTAAATAAATATCAAGAGCCTGATGTTGTTAACTTTGATTTTAAAA
ACATGGCTGATGGACAACTTCACCACATAATGATTAACAGAGAAGAAGGAGTGGTCTT
TATAGAGATTGACGATAATAGAAGGAGACAAGTTCACCTGTCATCAGGCACAGAATTC
AGTGCAGTCAAATCTCTGGTATTGGGCAGGATTTTAGAACACAGTGATGTGGACCAGG
AGACTGCACTGGCAGGTGCGCAGGGCTTCACAGGCTGCCTCTCTGCAGTGCAGCTCAG
CCACGTGGCCCCTCTGAAGGCAGCTCTGCACCCCAGCCACCCAGACCCTGTCACTGTT
ACAGGACACGTGACCGAGTCCAGCTGTATGGCCCAGCCTGGCACTGATGCCACATCAA
GGGAAAGGACACACTCGTTTGCAGATCATTCTGGAACAATAGATGACAGAGAGCCCCT
TGCTAATGCAATCAAAAGTGACTCTGCAGTAATTGGAGGTCTGATAGCTGTTGTGATT
TTTATCTTGCTTTGCATCACTGCCATAGCTGTTCGCATTTATCAGCAGAAAAGGTTAT
ATAAAAGAAGTGAGGCAAAAAGGTCAGAGAATGTAGACAGTGCTGAGGCTGTTCTGAA
AAGTGAGCTTAATATACAAAATGCAGTCAATGAAAATCAGAAAGAGTACTTCTTC**TGA**
TTGGCAGCTATGATTTAACATAAAATTATGATAGTTTGTTTTAATAGCCAGGGGTTCT
CAATGGAAAAACGAATGCTCTTACACTGAATGTACAGGCAGTGGGCTTGCAGCACTGC
CATCTTGCCATGTACAGGCTTGGGGTGGCTCCAGGAAGCCTCGTCCAGTGATATATTT
CTCATAGCATTCATTCTATGGAACAAGAAATTAGATATTGCTGTTAATTTTCAACTGT
TCTGGTATGATCTAAAACAAGTTTAACCTGCTTAATGGCTACAGTTTTTACATGTGAA
AACTGTAGCCTTGGTCTCTTAACCATGTAATACATAAGTTTTGTTAGAGGTAAAAATT
AAATTTGGACTATAATGTCCTTGCTTTATTTG

	ORF Start: ATG at 18	ORF Stop: TGA at 3942
	SEQ ID NO: 20	1308 aa	MW at 145314.9kD
NOV7a, CG58586-01 Protein Sequence	MGSVTGAVLKTLLLLSTQNWNRVEAGNSYDCDDPLVSALPQASFSSSSELSSSHGPGF ARLNRRDGAGGWSPLVSNKYQWLQIDLGERMEVTAVATQGGYGSSNWVTSYLLMFSDS GWNWKQYRQEDSIWGFSGNANADSVVYYRLQPSIKARFLRFIPLEWNPKGRIGMRIEV FGCAYRSEVVDLDGKSSLLYRFDQKSLSPIKDIISLKFKTMQSDGILLHREGPNGDHI TLQLRRARLFLLINSGEAKLPSTSTLVNLTLGSLLDDQHWHSVLIQRLGKQVNFTVDE HRHHFHARGEFNLMNLDYEISFGGIPAPGKSVSFPHRNFHGCLENLYYNGVDIIDLAK QQKPQIIAMGNVSFSCSQPQSMPVTFLSSRSYLALPDFSGEEEVSATFQFRTWNKAGL LLFSELQLISGGILLFLSDGKLKSNLYQPGKLPSDITAGVELNDGQWHSVSLSAKKNH LSVAVDGQMASAAPLLGPEQIYSGGTYYFGGCPDKSFGSKCKSPLGGFQGCMRLISIS GKVVDLISVQQGSLGNFSDLQIDSCGISDRCLPNYCEHGGECSQSWSTFHCNCTNTGY RGATCHNSIYEQSCEAYKHRGNTSGFYYIDSDGSGPLEPFLLYCNMTETAWTIIQHNG SDLTRVRNTNPENPYAGFFEYVASMEQLQATINRAEHCEQEFTYYCKKSRLVNKQDGT PLSWWVGRTNETQTYWGGSSPDLQKCTCGLEGNCIDSQYYCNCDADRNEWTNDTGLLA YKEHLPVTKIVITDTGRLHSEAAYKLGPLLCRGDRSFWNSASFDTEASYLHFPTFHGE LSADVSFFFKTTASSGVFLENLGIADFIRIELRSPTVVTFSFDVGNGPFEISVQSPTH FNDNQWHHVRVERNMKEASLQVDQLTPKTQPAPADGHVLLQLNSQLFVGGTATRQRGF LGCIRSLQLNGMTLDLEERAQVTPEVQPGCRGHCSSYGKLCRNGGKCRERPIGFFCDC TFSAYTGPFCSNEISAYFGSGSSVIYNFQENYLLSKNSSSHAASFHGDMKLSREMIKF SFRTTRTPSLLLFVSSFYKEYLSVIIAKNGSLQIRYKLNKYQEPDVVNFDFKNMADGQ LHHIMINREEGVVFIEIDDNRRRQVHLSSGTEFSAVKSLVLGRILEHSDVDQETALAG AQGFTGCLSAVQLSHVAPLKAALHPSHPDPVTVTGHVTESSCMAQPGTDATSRERTHS FADHSGTIDDREPLANAIKSDSAVIGGLIAVVIFILLCITAIAVRIYQQKRLYKRSEA KRSENVDSAEAVLKSELNIQNAVNENQKEYFF
	SEQ ID NO: 21	4331 bp
NOV7b, CG58586-02 DNA Sequence	GCCCAATAAATGTGAAC**ATG**GGATCTGTCACGGGAGCTGTCCTCAAGACGCTACTTCT GTTATCTACTCAAAATTGGAACAGAGTCGAAGCTGGGAATTCCTATGACTGTGATGAT CCTCTTGTGTCTGCCTTGCCTCAGGCATCCTTCAGCAGTTCTTCCGAGCTCTCCAGCA GTCATGGTCCTGGATTTGCAAGGCTGAATAGAAGAGATGGAGCTGGTGGCTGGTCTCC ACTTGTGTCTAACAAATACCAGTGGTTGCAGATTGACCTTGGAGAGAGAATGGAGGTC ACCGCTGTGGCCACTCAAGGGGGATATGGTAGCTCCAACTGGGTGACCAGCTACCTCC

```
TGATGTTCAGTGATAGTGGCTGGAACTGGAAACAATATCGCCAAGAGGACAGCATCTG
GGGTTTTTCAGGAAATGCAAATGCAGACAGTGTTGTGTACTATAGACTCCAGCCTTCT
ATCAAAGCCAGATTTCTGCGCTTCATCCCTTTGGAATGGAACCCCAAGGGCAGAATTG
GAATGCGAATCGAAGTGTTCGGATGTGCATACAGATCAGAAGTGGTTGATCTTGATGG
AAAAAGTTCCCTTCTCTACAGATTTGATCAAAAATCCCTGAGCCCAATAAAAGACATT
ATTTCTTTGAAATTCAAAACCATGCAGAGTGATGGGATTCTACTCCACAGGGAAGGGC
CAAATGGAGATCACATCACACTGCAATTAAGAAGAGCAAGACTCTTTTTACTTATTAA
TTCAGGTGAAGCTAAACTGCCTTCCACTTCCACCCTGGTCAATCTCACCCTGGGCAGC
CTGCTAGATGATCAGCATTGGCATTCAGTGCTCATCCAGCGTTTGGGCAAACAAGTCA
ACTTCACAGTGGACGAACACAGGCATCATTTCCATGCACGGGGAGAATTCAATCTCAT
GAATCTTGATTATGAGATCAGCTTTGGAGGGATTCCAGCACCTGGAAAATCAGTGTCA
TTCCCACATAGAAATTTTCATGGATGTTTAGAAAATCTCTATTATAATGGAGTGGATA
TCATTGATTTGGCCAAGCAGCAAAAACCACAGATCATTGCTATGGGAAATGTGTCATT
TTCTTGTTCACAACCACAATCTATGCCCGTGACTTTTCTGAGCTCCAGGAGTTATTTA
GCACTGCCAGACTTCTCTGGAGAGGAGGAGGTTTCTGCCACTTTTCAATTTCGAACTT
GGAATAAGGCAGGGCTTCTGCTGTTCAGTGAACTTCAGCTGATTTCAGGGGGTATCCT
CCTCTTTCTGAGTGATGGAAAACTTAAGTCGAATCTCTACCAGCCAGGAAAATTACCC
AGTGACATCACAGCAGGTGTCGAATTAAATGATGGGCAGTGGCATTCTGTCTCTTTAT
CTGCTAAAAAGAATCACTTGAGTGTGGCGGTGGACGGCCAGATGGCTTCTGCTGCTCC
TCTGCTGGGGCCTGAGCAGATTTATTCGGGTGGCACCTATTATTTTGGAGGTTGTCCT
GACAAAAGCTTTGGATCCAAATGTAAAAGTCCACTTGGTGGATTTCAGGGATGTATGA
GGCTCATTTCTATCAGCGGCAAAGTGGTAGATCTGATTTCAGTTCAGCAGGGGTCCCT
TGGGAACTTCAGTGACCTTCAGATAGACTCATGTGGCATCTCAGACAGGTGTTTGCCC
AACTATTGTGAACACGGTGGGGAGTGTTCCCAGTCCTGGAGCACCTTTCATTGTAACT
GTACCAACACTGGTTACAGAGGAGCTACTTGCCATAACGCTATCTATGAGCAGTCATG
TGAAGCCTATAAGCACAGAGGAAATACTTCAGGGTTTTACTATATAGATTCAGATGGA
AGTGGTCCCCTGGAACCATTTCTTCTATATTGCAATATGACCCAAGAAACTGCATGGA
CCATCATACAGCACAACGGCTCTGACTTAACAAGAGTCAGAAATACTAATCCAGAGAA
CCCATATGCTGGGTTTTTCGAGTATGTGGCCAGCATGGAGCAACTTCAGGCCACTATT
AACCGTGCAGAGCACTGTGAACAGGAGTTTACTTATTACTGCAAGAAGTCACGGCTGG
TCAATAAGCAAGATGGAACCCCTCTGAGTTGGTGGGTAGGAAGAACCAATGAAACGCA
AACCTACTGGGGAGGTTCTTCGCCTGATCTTCAAAAATGTACTTGTGGATTAGAGGGA
AACTGCATTGATTCTCAGTATTACTGCAATTGTGATGCTGACCGGAATGAATGGACCA
ATGACACTGGATTGCTTGCTTATAAAGAACATCTTCCAGTAACTAAGATCGTGATTAC
AGACACAGGCCGACTGCATTCAGAAGCAGCTTATAAACTGGGCCTCTGCTCTGCCGG
GGAGACAGTAAGTGGTCATTTTGGAATTCAGCTTCCTTTGATACCGAGGCTTCATATC
TTCATTTTCCTACCTTCCACGGAGAACTTAGCGCGGATGTATCTTTCTTTTTTAAGAC
AACAGCTTCATCTGGGGTATTTTTAGAGAACTTGGGGATTGCTGATTTTATACGGATA
GAGCTTCGCACAGTAGTGACTTTTTCATTTGATGTGGGGAATGGGCCTTTTGAAATCT
CAGTGCAGTCACCCACCCACTTCAACGACAACCAGTGGCACCATGTGAGGGGTTGAAAG
GAACATGAAGGAGGCCTCCCTTCAAGTGGATCAGCTGACACCAAAGACACAGCCCGCC
CCCGCTGATGGGCATGTCCTGTTACAGCTCAACAGTCAGCTCTTCGTGGGTGGAACGG
CCACCAGACAGAGAGGCTTTCTGGGCTGCATTCGGTCTCTGCAGTTGAATGGGATGAC
CCTGGATTTGGAAGAAAGAGCCCAGGTGACTCCAGAAGTGCAGCCAGGTTGTAGGGGA
CATTGCAGCAGCTATGGGAAGTTATGCCGCAATGGAGGGAAATGCAGAGAAAGACCCA
TTGGGTTCTTTTGTGACTGCACTTTCTCTGCATACACAGGGCCATTCTGCTCAAATGA
GATTTCTGCATATTTTGGATCTGGCTCATCCGTGATATACAATTTTCAAGAAAATTAT
CTTTTAAGTAAAAACTCCAGCTCCCACGCTGCTTCATTTCATGGTGATATGAAGCTGA
GCAGAGAAATGATCAAATTTAGTTTCCGAACAACACGAACACCAAGCTTGCTGCTTTT
TGTGAGCTCCTTTTACAAAGAATACCTTTCTGTGATCATTGCCAAAAATGGAAGTTTG
CAGATCAGGTACAAGTTAAATAAATATCAAGAGCCTGATGTTGTTAACTTTGATTTTA
AAAACATGGCTGATGGACAACTTCACCACATAATGATTAACAGAGAAGAAGGAGTGGT
CTTTATAGAGATTGACGATAATAGAAGGAGACAAGTTCACCTGTCATCAGGCACAGAA
TTCAGTGCAGTCAAATCTCTGGTATTGGGCAGGATTTTAGAACACAGTGATGTGGACC
AGGAGACTGCACTGGCAGGTGCGCAGGGCTTCACAGGCTGCCTCTCTGCAGTGCAGCT
CAGCCACGTGGCCCCTCTGAAGGCAGCTCTGCACCCCAGCCACCCAGACCCTGTCACT
GTTACAGGACACGTGACCGAGTCCAGCTGTATGGCCCAGCCTGGCACTGATGCCACAT
CAAGGGAAAGGACACACTCGTTTGCAGATCATTCTGGAACAATAGATGACAGAGAGCC
CCTTGCTAATGCAATCAAAAGTGACTCTGCAGTAATTGGAGGTCTGATAGCTGTTGTG
```

	ATTTTTATCTTGCTTTGCATCACTGCCATAGCTGTTCGCATTTATCAGCAGAAAAGGT TATATAAAAGAAGTGAGGCAAAAAGGTCAGAGAATGTAGACAGTGCTGAGGCTGTTCT GAAAAGTGAGCTTAATATACAAAATGCAGTCAATGAAAATCAGAAAGAGTACTTCTTC TGATTGGCAGCTATGATTTAACATAAAATTATGATAGTTTGTTTTAATAGCCAGGGGT TCTCAATGGAAAAACGAATGCTCTTACACTGAATGTACAGGCAGTGGGCTTGCAGCAC TGCCATCTTGCCATGTACAGGCTTGGGGTGGCTCCAGGAAGCCTCGTCCAGTGATATA TTTCTCATAGCATTCATTCTATGGAACAAGAAATTAGATATTGCTGTTAATTTTCAAC TGTTCTGGTATGATCTAAAACAAGTTTAACCTGCTTAATGGCTACAGTTTTTACATGT GAAAACTGTAGCCTTGGTCTCTTAACCATGTAATACATAAGTTTTGTTAGAGGTAAAA ATTAAATTTGGACTATAATGTCCTTGCTTTATTTGNNNN
	ORF Start: ATG at 18 ‖ ORF Stop: TGA at 3945
	SEQ ID NO: 22 ‖ 1309 aa ‖ MW at 145488.1kD
NOV7b, CG58586-02 Protein Sequence	MGSVTGAVLKTLLLLSTQNWNRVEAGNSYDCDDPLVSALPQASFSSSSELSSSHGPGF ARLNRRDGAGGWSPLVSNKYQWLQIDLGERMEVTAVATQGGYGSSNWVTSYLLMFSDS GWNWKQYRQEDSIWGFSGNANADSVVYYRLQPSIKARFLRFIPLEWNPKGRIGMRIEV FGCAYRSEVVDLDGKSSLLYRFDQKSLSPIKDIISLKFKTMQSDGILLHREGPNGDHI TLQLRRARLFLLINSGEAKLPSTSTLVNLTLGSLLDDQHWHSVLIQRLGKQVNFTVDE HRHHFHARGEFNLMNLDYEISFGGIPAPGKSVSFPHRNFHGCLENLYYNGVDIIDLAK QQKPQIIAMGNVSFSCSQPQSMPVTFLSSRSYLALPDFSGEEEVSATFQFRTWNKAGL LLFSELQLISGGILLFLSDGKLKSNLYQPGKLPSDITAGVELNDGQWHSVSLSAKKNH LSVAVDGQMASAAPLLGPEQIYSGGTYYFGGCPDKSFGSKCKSPLGGFQGCMRLISIS GKVVDLISVQQGSLGNFSDLQIDSCGISDRCLPNYCEHGGECSQSWSTFHCNCTNTGY RGATCHNAIYEQSCEAYKHRGNTSGFYYIDSDGSGPLEPFLLYCNMTQETAWTIIQHN GSDLTRVRNTNPENPYAGFFEYVASMEQLQATINRAEHCEQEFTYYCKKSRLVNKQDG TPLSWWVGRTNETQTYWGGSSPDLQKCTCGLEGNCIDSQYYCNCDADRNEWTNDTGLL AYKEHLPVTKIVITDTGRLHSEAAYKLGPLLCRGDSKWSFWNSASFDTEASYLHFPTF HGELSADVSFFFKTTASSGVFLENLGIADFIRIELRTVVTFSFDVGNGPFEISVQSPT HFNDNQWHHVRVERNMKEASLQVDQLTPKTQPAPADGHVLLQLNSQLFVGGTATRQRG FLGCIRSLQLNGMTLDLEERAQVTPEVQPGCRGHCSSYGKLCRNGGKCRERPIGFFCD CTFSAYTGPFCSNEISAYFGSGSSVIYNFQENYLLSKNSSHAASFHGDMKLSREMIK FSFRTTRTPSLLLFVSSFYKEYLSVIIAKNGSLQIRYKLNKYQEPDVVNFDFKNMADG QLHHIMINREEGVVFIEIDDNRRRQVHLSSGTEFSAVKSLVLGRILEHSDVDQETALA GAQGFTGCLSAVQLSHVAPLKAALHPSHPDPVTVTGHVTESSCMAQPGTDATSRERTH SFADHSGTIDDREPLANAIKSDSAVIGGLIAVVIFILLCITAIAVRIYQQKRLYKRSE AKRSENVDSAEAVLKSELNIQNAVNENQKEYFF

Sequence comparison of the above protein sequences yields the following

sequence relationships shown in Table 7B.

Table 7B. Comparison of NOV7a against NOV7b.		
Protein Sequence	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV7b	1..1308 1..1309	1293/1311 (98%) 1294/1311 (98%)

Further analysis of the NOV7a protein yielded the following properties shown in
Table 7C.

Table 7C. Protein Sequence Properties NOV7a	
PSort analysis:	0.8343 probability located in mitochondrial inner membrane; 0.6400 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 26 and 27

A search of the NOV7a protein against the Geneseq database, a proprietary
database that contains sequences published in patents and patent publication, yielded
5   several homologous proteins shown in Table 7D.

Table 7D. Geneseq Results for NOV7a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE07282	Human neurexin-like protein #1 - Homo sapiens, 1307 aa. [WO200158938-A2, 16-AUG-2001]	12..1307 11..1306	761/1299 (58%) 981/1299 (74%)	0.0
AAE07293	Human neurexin-like protein #12 - Homo sapiens, 1298 aa. [WO200158938-A2, 16-AUG-2001]	30..1307 20..1297	758/1281 (59%) 975/1281 (75%)	0.0
AAE07294	Human neurexin-like protein #13 - Homo sapiens, 1175 aa. [WO200158938-A2, 16-AUG-2001]	137..1307 2..1174	686/1174 (58%) 890/1174 (75%)	0.0
AAB42887	Human ORFX ORF2651 polypeptide sequence SEQ ID NO:5302 - Homo sapiens, 1339 aa. [WO200058473-A2, 05-OCT-2000]	28..1306 40..1337	634/1302 (48%) 873/1302 (66%)	0.0
AAM41859	Human polypeptide SEQ ID NO 6790 - Homo sapiens, 1355 aa. [WO200153312-A1, 26-JUL-2001]	28..1252 40..1281	618/1246 (49%) 838/1246 (66%)	0.0

In a BLAST search of public sequence databases, the NOV7a protein was found to
have homology to the proteins shown in the BLASTP data in Table 7E.

Table 7E. Public BLASTP Results for NOV7a				
Protein Accession Number	Protein/Organism/Length	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9C0A0	Contactin associated protein-like 4 precursor (Cell recognition molecule Caspr4) - Homo sapiens (Human), 1308 aa.	1..1308 1..1308	1308/1308 (100%) 1308/1308 (100%)	0.0
Q8WX98	CELL RECOGNITION PROTEIN CASPR4 - Homo sapiens (Human), 1311 aa.	30..1308 33..1311	1279/1279 (100%) 1279/1279 (100%)	0.0
Q99P47	Contactin associated protein-like 4 precursor (Cell recognition molecule Caspr4) - Mus musculus (Mouse), 1310 aa.	1..1308 3..1310	1132/1308 (86%) 1210/1308 (91%)	0.0
AAG52889	CELL RECOGNITION MOLECULE CASPR3 - Homo sapiens (Human), 1288 aa.	1..1280 1..1284	905/1286 (70%) 1049/1286 (81%)	0.0
Q8WYK1	CASPR5 - Homo sapiens (Human), 1306 aa.	12..1307 11..1305	763/1299 (58%) 982/1299 (74%)	0.0

PFam analysis predicts that the NOV7a protein contains the domains shown in the Table 7F.

Table 7F. Domain Analysis of NOV7a			
Pfam Domain	**NOV7a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
F5_F8_type_C	34..174	60/161 (37%) 117/161 (73%)	3e-49
laminin_G	212..344	36/166 (22%) 94/166 (57%)	0.013
laminin_G	398..527	46/162 (28%) 91/162 (56%)	7.2e-16
EGF	553..585	12/47 (26%) 25/47 (53%)	0.0012
laminin_G	821..943	46/162 (28%) 97/162 (60%)	1.6e-20
EGF	962..996	14/47 (30%) 27/47 (57%)	0.089
laminin_G	1073..1131	14/68 (21%) 45/68 (66%)	0.37
TSPN	981..1178	33/229 (14%) 123/229 (54%)	0.71

EXAMPLE 8.

The NOV8 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 8A.

<table>
<tr><td colspan="3" align="center">Table 8A. NOV8 Sequence Analysis</td></tr>
<tr><td></td><td>SEQ ID NO: 23</td><td>5878 bp</td></tr>
<tr><td>NOV8a,
CG93453-01
DNA Sequence</td><td colspan="2">TCTTAAAGAAACTTATTTTGGGCGGGGGGGGTGGGTTTGCTCTGGGCATTTGCTTTGC
CCAGTAGTTGGAAAGTGAACTCGACTCGTGATGGTTCTCCTGTCACTTTGGTTGATAG
CAGCCGCTCTGGTAGAGGTTAGGACTTCAGCTGATGGACAAGCTGGTAATGAAGAAAT
GGTGCAAATAGATTTACCAATAAAGAGATATAGAGAGTATGAGCTGGTGACTCCAGTC
AGCACAAATCTAGAAGGACGCTATCTCTCCCATACTCTTTCTGCGAGTCACAAAAAGA
GGTCAGCGAGGGACGTGTCTTCCAACCCTGAGCAGTTGTTCTTTAACATCACGGCATT
TGGAAAAGATTTTCATCTGCGACTAAAGCCCAACACTCAACTAGTAGCTCCTGGGGCT
GTTGTGGAGTGGCATGAGACATCTCTGGTGCCTGGGAATATAACCGATCCCATTAACA
ACCATCAACCAGGAAGTGCTACGTATAGAATCCGGAGAACAGAGCCTTTGCAGACTAA
CTGTGCTTATGTTGGTGACATCGTGGACATTCCAGGAACCTCTGTTGCCATCAGCAAC
TGTGATGGTCTGGCTGGAATGATAAAAAGTGATAATGAAGAGTATTTCATTGAACCCT
TGGAAGAGGTAAACAGATGGAGGAAGAAAAAGGAAGGATTCATGTTGTCTACAAGAG
ATCAGCTGTAGAACAGGCTCCCATAGACATGTCCAAAGACTTCCACTACAGAGAGTCG
GACCTGGAAGGCCTTGATGATCTAGGTACTGTTTATGGCAACATCCACCAGCAGCTGA
ATGAAACAATGAGACGCCGCAGACACGCGGGAGAAAACGATTACAATATCGAGGTACT
GCTGGGAGTGGATGACTCTGTGGTCCGTTTCCATGGCAAAGAGCACGTCCAAAACTAC
CTCCTGACCCTAATGAACATTGTGAATGAAATTTACCATGATGAGTCCCTCGGAGTGC
ATATAAATGTGGTCCTGGTGCGCATGATAATGCTGGGATATGCAAAGTCCATCAGCCT
CATAGAAAGGGGAAACCCATCCAGAAGCTTGGAGAATGTGTGTCGCTGGGCGTCCCAA
CAGCAAAGATCTGATCTCAACCACTCTGAACACCATGACCATGCAATTTTTTTAACCA
GGCAAGACTTTGGACCTGCTGGAATGCAAGGATATGCTCCAGTCACCGGCATGTGTCA
TCCAGTGAGAAGTTGTACCCTGAATCATGAGGATGGTTTTTCATCTGCTTTTGTAGTA
GCCCATGAAACGGGCCATGTGTTGGGAATGGAGCATGATGGACAAGGCAACAGGTGTG
GTGATGAGACTGCTATGGGAAGTGTCATGGCTCCCTTGGTACAAGCAGCATTCCATCG
TTACCACTGGTCCCGATGCAGTGGTCAAGAACTGAAAAGATATATCCATTCCTATGAC
TGTCTCCTTGATGACCCTTTTGATCATGATTGGCCTAAACTCCCAGAACTTCCTGGAA
TCAATTATTCTATGGATGAGCAATGTCGTTTTGATTTTGGTGTTGGCTATAAAATGTG
CACCGCGTTCCGAACCTTTGACCCATGTAAACAGCTGTGGTGGTAGCCATCCTGATAAT
CCCTACTTTTGTAAGACTAAAAAGGGACCTCCACTTGATGGGACTGAATGTGCTGCTG
GAAAATGGTGCTATAAGGGTCATTGCATGTGGAAGAATGCTAATCAGCAAAAACAAGA
TGGCAATTGGGGGTCATGGACTAAATTTGGCTCCTGTTCTCGGACATGTGGAACTGGT
GTTCGTTTCAGAACACGCCAGTGCAATAATCCCATGCCCATCAATGGTGGTCAGGATT
GTCCTGGTGTTAATTTTGAGTACCAGCTTTGTAACACAGAAGAATGCCAAAAACACTT
TGAGGACTTCAGAGCACAGCAGTGTCAGCAGCGAAACTCCCACTTTGAATACCAGAAT
ACCAAACACCACTGGTTGCCATATGAACATCCTGACCCCAAGAAAAGATGCCACCTTT
ACTGTCAGTCCAAGGAGACTGGAGATGTTGCTTACATGAAACAACTGGTGCATGATGG
AACGCACTGTTCTTACAAAGATCCATATAGCATATGTGTGCGAGGAGAGTGTGTGAAA
GTGGGCTGTGATAAAGAAATTGGTTCTAATAAGGTTGAGGATAAGTGTGGTGTCTGTG
GAGGAGATAATTCCCACTGCCGAACCGTGAAGGGGACATTTACCAGAACTCCCAGGAA
GCTTGGGTACCTTAAGATGTTTGATATACCCCCTGGGGCTAGACATGTGTTAATCCAA
GAAGACGAGGCTTCTCCTCATATTCTTGCTATTAAGAACCAGGCTACAGGCCATTATA
TTTTAAATGGCAAAGGGGAGGAAGCCAAGTCGCGGACCTTCATAGATCTTGGTGTGGA
GTGGGATTATAACATTGAAGATGACATTGAAAGTCTTCACACCGATGGACCTTTACAT
GATCCTGTTATTGTTTTGATTATACCTCAAGAAAATGATACCCGCTCTAGCCTGACAT
ATAAGTACATCATCCATGAAGACTCTGTACCTACAATCAACAGCAACAATGTCATCCA
GGAAGAATTAGATACTTTTGAGTGGGCTTTGAAGAGCTGGTCTCAGTGTTCCAAACCC
TGTGGTGGAGGTTTCCAGTACACTAAATATGGATGCCGTAGGAAAGTGATAATAAAA
TGGTCCATCGCAGCTTCTGTGAGGCCAACAAAAAGCCGAAACCTATTAGACGAATGTG
CAATATTCAAGAGTGTACACATCCACTCTGGGTAGCAGAAGAATGGGAACACTGCACC
AAAACCTGTGGAAGTTCTGGCTATCAGCTTCGCACTGTACGCTGCCTTCAGCCACTCC</td></tr>
</table>

	TTGATGGCACCAACCGCTCTGTGCACAGCAAATACTGCATGGGTGACCGTCCCGAGAG CCGCCGGCCCTGTAACAGAGTGCCCTGCCCTGCACAGTGGAAAACAGGACCCTGGAGT GAGTGTTCAGTGACCTGCGGTGAAGGAACGGAGGTGAGGCAGGTCCTCTGCAGGGCTG GGGACCACTGTGATGGTGAAAAGCCTGAGTCGGTCAGAGCCTGTCAACTGCCTCCTTG TAATGATGAACCATGTTTGGGAGACAAGTCCATATTCTGTCAAATGGAAGTGTTGGCA CGATACTGCTCCATACCAGGTTATAACAAGTTATGTTGTGAGTCCTGCAGCAAGCGCA GTAGCACCCTGCCACCACCATACCTTCTAGAAGCTGCTGAAACTCATGATGATGTCAT CTCTAACCCTAGTGACCTCCCTAGATCTCTAGTGATGCCTACATCTTTGGTTCCTTAT CATTCAGAGACCCCTGCAAAGAAGATGTCTTTGAGTAGCATCTCTTCAGTGGGAGGTC CAAATGCATATGCTGCTTTCAGGCCAAACAGTAAACCTGATGGTGCTAATTTACGCCA GAGGAGTGCTCAGCAAGCAGGAAGTAAGACTGTGAGACTGGTCACCGTACCATCCTCC CCACCCACCAAGAGGGTCCACCTCAGTTCAGCTTCACAAATGGCTGCTGCTTCCTTCT TTGCAGCCAGTGATTCAATAGGTGCTTCTTCTCAGGCAAGAACCTCAAAGAAAGATGG AAAGATCATTGACAACAGACGTCCGACAAGATCATCCACCTTAGAAAGA<u>TGA</u>GAAAGT <u>GAACCAAAAAGGCTAGAAACCAGAGGAAAACCTGGACAACCTCTCTCTTCCCATGGTG CATATGCTTGTTTAAAGTGGAAATCTCTATAGATCGTCAGCTCATTTTATCTGTAATT GGAAGAACAGAAAGTGCTGGCTCACTTTCTAGTTGCTTTCATCCTCCTTTTGTTCTGC ATTGACTCATTTACCAGAATTCATTGGAAGAAATCACCAAAGATTATTACAAAAGAAA AATATGTTGCTAAGATTGTGTTGGTCGCTCTCTGAAGCAGAAAAGGGACTGGAACCAA TTGTGCATATCAGCTGACTTTTTGTTTGTTTTAGAAAAGTTACAGTAAAAATTAAAAA GAGATACCAATGGTTTACACTTTAACAAGAAATTTTGGATATGGAACAAAGAATTCTT AGACTTGTATTCCTATTTATCTATATTAGAAATATTGTATGAGCAAATTTGCAGCTGT TGTGTAAATACTGTATATTGCAAAAATCAGTATTATTTTAAGAGATGTGTTCTCAAAT GATTGTTTACTATATTACATTTCTGGATGTTCTAGGTGCCTGTCGTTGAGTATTGCCT TGTTTGACATTCTATAGGTTAATTTTCAAAGCAGAGTATTACAAAAGAGAAGTTAGAA TTACAGCTACTGACAATATAAAGGGTTTTGTTGAATCAACAATGTGATACGTAAATTA TAGAAAAAGAAAAGAAACACAAAAGCTATAGATATACAGATATCAGCTTACCTATTGC CTTCTATACTTATAATTTAAAGGATTGGTGTCTTAGTACACTTGTGGTCACAGGGATC AACGAATAGTAAATAATGAACTCGTGCAAGACAAAACTGAAACCCTCTTTCCAGGACC TCAGTAGGCACCGTTGAGGTGTCCTTTGTTTTTGTGTGTGTGTGTTCTTTTTTAATTT TCGCATTGTTGACAGATACAAACAGTTATACTCAATGTACTGTAATAATCGCAAAGGA AAAAGTTTTGGGATAACTTATTTGTATGTTGGTAGCTGAGAAAAATATCATCAGTCTA GAATTGATATTTGAGTATAGTAGAGCTTTGGGGCTTTGAAGGCAGGTTCAAGAAAGCA TATGTCGATGGTTGAGATATTTATTTTCCATATGGTTCATGTTCAAATGTTCACAACC ACAATGCATCTGACTGCAATAATGTGCTAATAATTTATGTCAGTAGTCACCTTGCTCA CAGCAAAGCCAGAAATGCTCTCTCCAGGGAGTAGATGTAAAGTACTTGTACATAGAAT TCAGAACTGAAGATATTTATTAAAAGTTGATTTTTTTTTCTTGATAGTATTTTTATGT ACTAAATATTTACACTAATATCAATTACATATTTTGGTAAACTAGAGAGACATAATTA GAGATGCATGCTTCGTTCTGTGCATAGAGACCTTTAAGCAAACTACTACAGCCAACTC AAAAGCTAAAACTGAACAAATTTGATGTTATACAAACATCTTGCATTTTTAGTAGTTG ATATTAAGTTGATGACTTGTTTCCCTTCAAGGAAACATTAAATTGTATGGACTCAGCT AGCTGTTCAATGAAATTGTGAATTAGAAACATTTTTAAAAGTTTTTGAAAGAGATAAG TGCATCATGAATTACATGTACATGAGAGGAGATAGTGATATCAGCATAATGATTTTGA GGTCAGTACCTGAGCTGTCTAAAAATATATTATACAAACTAAAATGTAGATGAATTAA CCTCTCAAAGCACAGAATGTGCAAGAACTTTTGCATTTTAATCGTTGTAAACTAACAG CTTAAACTATTGACTCTATACCTCTAAAGAATTGCTGCTACTTTGTGCAAGAACTTTG AAGGTCAAATTAGGCAAATTCCAGATAGTAAAACAATCCCTAAGCCTTAAGTCTTTTT TTTTTCCTAAAAATTCCCATAGAATAAAATTCTCTCTAGTTTACTTGTGTGTGCATAC ATCTCATCCACAGGGGAAGATAAAGATGGTCACACAAACAGTTTCCATAAAGATGTAC ATATTCATTATACTTCTGACCTTTGGGCTTTCTTTTCTACTAAGCTAAAAATTCCTTT TTATCAAAGTGTACACTACTGATGCTGTTTGTTGTACTGAGAGCACGTACCAATAAAA ATGTTAACAAAATATAAAAA</u>
	ORF Start: ATG at 89	ORF Stop: TGA at 3704
	SEQ ID NO: 24	1205 aa	MW at 135601.5kD
NOV8a, CG93453-01 Protein	MVLLSLWLIAAALVEVRTSADGQAGNEEMVQIDLPIKRYREYELVTPVSTNLEGRYLS HTLSASHKKRSARDVSSNPEQLFFNITAFGKDFHLRLKPNTQLVAPGAVVEWHETSLV PGNITDPINNHQPGSATYRIRRTEPLQTNCAYVGDIVDIPGTSVAISNCDGLAGMIKS DNEEYFIEPLERGKQMEEEKGRIHVVYKRSAVEQAPIDMSKDFHYRESDLEGLDDLGT

Sequence	VYGNIHQQLNETMRRRRHAGENDYNIEVLLGVDDSVVRFHGKEHVQNYLLTLMNIVNE
	IYHDESLGVHINVVLVRMIMLGYAKSISLIERGNPSRSLENVCRWASQQQRSDLNHSE
	HHDHAIFLTRQDFGPAGMQGYAPVTGMCHPVRSCTLNHEDGFSSAFVVAHETGHVLGM
	EHDGQGNRCGDETAMGSVMAPLVQAAFHRYHWSRCSGQELKRYIHSYDCLLDDPFDHD
	WPKLPELPGINYSMDEQCRFDFGVGYKMCTAFRTFDPCKQLWCSHPDNPYFCKTKKGP
	PLDGTECAAGKWCYKGHCMWKNANQQKQDGNWGSWTKFGSCSRTCGTGVRFRTRQCNN
	PMPINGGQDCPGVNFEYQLCNTEECQKHFEDFRAQQCQQRNSHFEYQNTKHHWLPYEH
	PDPKKRCHLYCQSKETGDVAYMKQLVHDGTHCSYKDPYSICVRGECVKVGCDKEIGSN
	KVEDKCGVCGGDNSHCRTVKGTFTRTPRKLGYLKMFDIPPGARHVLIQEDEASPHILA
	IKNQATGHYILNGKGEEAKSRTFIDLGVEWDYNIEDDIESLHTDGPLHDPVIVLIIPQ
	ENDTRSSLTYKYIIHEDSVPTINSNNVIQEELDTFEWALKSWSQCSKPCGGGFQYTKY
	GCRRKSDNKMVHRSFCEANKKPKPIRRMCNIQECTHPLWVAEEWEHCTKTCGSSGYQL
	RTVRCLQPLLDGTNRSVHSKYCMGDRPESRRPCNRVPCPAQWKTGPWSECSVTCGEGT
	EVRQVLCRAGDHCDGEKPESVRACQLPPCNDEPCLGDKSIFCQMEVLARYCSIPGYNK
	LCCESCSKRSSTLPPPYLLEAAETHDDVISNPSDLPRSLVMPTSLVPYHSETPAKKMS
	LSSISSVGGPNAYAAFRPNSKPDGANLRQRSAQQAGSKTVRLVTVPSSPPTKRVHLSS
	ASQMAAASFFAASDSIGASSQARTSKKDGKIIDNRRPTRSSTLER
	SEQ ID NO: 25 2286 bp
NOV8b, CG93453-02 DNA Sequence	CTCGAGACAAGGAGGCAGTTGACAGGCTCTGACCGACTCAGGCTTTTCACCATCACAG
	TGGTCCCCAGCCCTGCAGAGGACCTGCCTCACCTCCGTTCCTTCACCGCAGGTCACTG
	AACACTCACTCCAGGGTCCTGTTTTCCACTGTGCAGGGCAGGGCACTCTGTTACAGGG
	CCGGCGGCTCTCGGGACGGTCACCCATGCAGTATTTGCTGTGCACAGAGCGGTTGGTG
	CCATCAAGGAGTGGCTGAAGGCAGCGTACAGTGCGAAGCTGATAGCCAGAACTTCCAC
	AGGTTTTGGTGCAGTGTTCCCATTCTTCTGCTACCCAGAGTGGATGTGTACACTCTTG
	AATATTGCACATTCGTCTAATAGGTTTCGGCTTTTTGTTGGCCTCACAGAAGCTGCGA
	TGGACCATTTTATTATCACTTTTCCTACGGCATCCATATTTAGTGTACTGGAAACCTC
	CACCACAGGGTTTGGAACACTGAGACCAGCTCTTCAAAGCCCACTCAAAAGTATCTAA
	TTCTTCCTGGATGACATTGTTGCTGTTGATTGTAGGTACAGAGTCTTCATGGATGATG
	TACTTATATGTCAGGCTAGAGCGGGTATCATTTTCTTGAGGTATAATCAAAACAATAA
	CAGGATCATGTAAAGGTCCATCGGTGTGAAGACTTTCAATGTCATCTTCAATGTTATA
	ATCCCACTCCACACCAAGATCTATGAAGGTCCGCGACTTGGCTTCCTCCCCTTTGCCA
	TTTAAAATATAATGGCCTGTAGCCTGGTTCTTAATAGCAAGAATATGAGGAGAAGCCT
	CGTCTTCTTGGATTAA**CAC**ATGTCTAGCCCCAGGGGGTATATCAAACATCTTAAGGTA
	CCCAAGCTTCCTGGGAGTTCTGGTAAATGTCCCCTTCACGGTTCGGCAGTGGGAATTA
	TCTCCTCCACAGACACCACACTTATCCTCAACCTTATTAGAACCAATTTCTTTATCAC
	AGCCCACTTTCACACACTCTCCTCGCACACATATGCTATATGGATCTTTGTAAGAACA
	GTGCGTTCCATCATGCACCAGTTGTTTCATGTAAGCAACATCTCCAGTCTCCTTGGAC
	TGACAGTAAAGGTGGCATCTTTTCTTGGGGTCAGGATGTTCATATGGCAACCAGTGGT
	GTTTGGTATTCTGGTATTCAAAGTGGGAGTTTCGCTGCTGACACTGCTGTGCTCTGAA
	GTCCTCAAAGTGTTTTTGGCATTCTTCTGTGTTACAAAGCTGGTACTCAAAATTAACA
	CCAGGACAATCCTGACCACCATTGATGGGCATGGGATTATTGCACTGGCGTGTCCTGA
	AACGAACACCAGTTCCACATGTCCGAGAACAGGAGCCAAATTTAGTCCATGACCCCCA
	ATTGCCATCTTGTTTTTGCTGATTAGCATTCTTCCACATGCAATGACCCTTATAGCAC
	CATTTTCCAGCAGCACATTCAGTCCCATCAAGTGGAGGTCCCTTTTTAGTCTTACAAA
	AGTAGGGATTATCAGGATGGCTACACCACAGCTGTTTACATGGGTCAAAGGTTCGGAA
	CGCGGTGCACATTTTATAGCCAACACCAAAATCAAAACGACATTGCTCATCCATAGAA
	TAATTGATTCCAGGAAGTTCTGGGAGTTTAGGCCAATCATGATCAAAAGGGTCATCAA
	GGAGACAGTCATAGGAATGGATATATCTTTTCAGTTCTTGACCACTGCATCGGGACCA
	GTGGTAACGATGGAATGCTGCTTGTACCAAGGGAGCCATGACACTTCCCATAGCAGTC
	TCATCACCACACCTGTTGCCTTGTCCATCATGCTCCATTCCCAACACATGGCCCGTTT
	CATGGGCTACTACAAAAGCAGATGAAAAACCATCCTCATGATTCAGGGTACAACTTCT
	CACTGGATGACACATGCCGGTGACTGGAGCATATCCTTGCATTCCAGCAGGTCCAAAG
	TCTTGCCTGGTTAAAAAATTGCATGGTCATGGTGTTCAGAGTGGTTGAGATCAGATC
	TTTGCTGTTGGGACGCCCAGCGACACACATTCTCCAAGCTTCTGGATGGGTTTCCCCT
	TTCTATGAGGCTGATGGACTTTGCATATCCCAGCATTATCATGCGCACCAGGACCACA
	TTTATATGCACTCCGAGGGACTCATCATGGTAAATTTCATTCACAATGTTCATTAGGG
	TCAGGAGGTAGTTTTGGACGTGCTCTTTGCCATGGAAACGGACCACAGAGTCATCCAC
	TCCCAGCAGTACCTCGATGGATCC

	ORF Start: at 829	ORF Stop: end of sequence	
	SEQ ID NO: 26	758 aa	MW at 86176.4kD
NOV8b, CG93453-02 Protein Sequence	IEVLLGVDDSVVRFHGKEHVQNYLLTLMNIVNEIYHDESLGVHINVVLVRMIMLGYAK SISLIERGNPSRSLENVCRWASQQQRSDLNHSEHHDHAIFLTRQDFGPAGMQGYAPVT GMCHPVRSCTLNHEDGFSSAFVVAHETGHVLGMEHDGQGNRCGDETAMGSVMAPNQQA AFHRYHWSRCSGQELKRYIHSYDCLLDDPFDHDWPKLPELPGINYSMDEQCRFDFGVG YKMCTAFRTFDPCKQLWCSHPDNPYFCKTKKGPPLDGTECAAGKWCYKGHCMWKNANQ QKQDGNWGSWTKFGSCSRTCGTGVRFRTRQCNNPMPINGGQDCPGVNFEYQLCNTEEC QKHFEDFRAQQCQQRNSHFEYQNTKHHWLPYEHPDPKKRCHLYCQSKETGDVAYMKQL VHDGTHCSYKDPYSICVRGECVKVGCDKEIGSNKVEDKCGVCGGDNSHCRTVKGTFTR TPRKLGYLKMFDIPPGARHVLIQEDEASPHILAIKNQATGHYILNGKGEEAKSRTFID LGVEWDYNIEDDIESLHTDGPLHDPVIVLIIPQENDTRSSLTYKYIIHEDSVPTINSN NVIQEELDTFEWALKSWSQCSKPCGGGFQYTKYGCRRKSDNKMVHRSFCEANKKPKPI RRMCNIQECTHPLWVAEEWEHCTKTCGSSGYQLRTVRCLQPLLDGTNRSVHSKYCMGD RPESRRGCNRVGCRAQWKTAGRSECSVTCGEGTEVRQVLCRAGDHCDGEKPESVRACQ LPGC
	SEQ ID NO: 27	2286 bp	
NOV8c, 210387874 DNA Sequence	GGATCCATCGAGGTACTGCTGGGAGTGGATGACTCTGTGGTCCGTTTCCATGGCAAAG AGCACGTCCAAAACTACCTCCTGACCCTAATGAACATTGTGAATGAAATTTACCATGA TGAGTCCCTCGGAGTGCATATAAATGTGGTCCTGGTGCGCATGATAATGCTGGGATAT GCAAAGTCCATCAGCCTCATAGAAAGGGGAAACCCATCCAGAAGCTTGGAGAATGTGT GTCGCTGGGCGTCCCAACAGCAAAGATCTGATCTCAACCACTCTGAACACCATGACCA TGCAATTTTTTTAACCAGGCAAGACTTTGGACCTGCTGGAATGCAAGGATATGCTCCA GTCACCGGCATGTGTCATCCAGTGAGAAGTTGTACCCTGAATCATGAGGATGGTTTTT CATCTGCTTTTGTAGTAGCCCATGAAACGGGCCATGTGTTGGGAATGGAGCATGATGG ACAAGGCAACAGGTGTGGTGATGAGACTGCTATGGGAAGTGTCATGGCTCCCTTGGTA CAAGCAGCATTCCATCGTTACCACTGGTCCCGATGCAGTGGTCAAGAACTGAAAAGAT ATATCCATTCCTATGACTGTCTCCTTGATGACCCTTTTGATCATGATTGGCCTAAACT CCCAGAACTTCCTGGAATCAATTATTCTATGGATGAGCAATGTCGTTTTGATTTTGGT GTTGGCTATAAAATGTGCACCGCGTTCCGAACCTTTGACCCATGTAAACAGCTGTGGT GTAGCCATCCTGATAATCCCTACTTTTGTAAGACTAAAAAGGGACCTCCACTTGATGG GACTGAATGTGCTGCTGGAAAATGGTGCTATAAGGGTCATTGCATGTGGAAGAATGCT AATCAGCAAAAACAAGATGGCAATTGGGGGTCATGGACTAAATTTGGCTCCTGTTCTC GGACATGTGGAACTGGTGTTCGTTTCAGGACACGCCAGTGCAATAATCCCATGCCCAT CAATGGTGGTCAGGATTGTCCTGGTGTTAATTTTGAGTACCAGCTTTGTAACACAGAA GAATGCCAAAAACACTTTGAGGACTTCAGAGCACAGCAGTGTCAGCAGCGAAACTCCC ACTTTGAATACCAGAATACCAAACACCACTGGTTGCCATATGAACATCCTGACCCCAA GAAAAGATGCCACCTTTACTGTCAGTCCAAGGAGACTGGAGATGTTGCTTACATGAAA CAACTGGTGCATGATGGAACGCACTGTTCTTACAAAGATCCATATAGCATATGTGTGC GAGGAGAGTGTGTGAAAGTGGGCTGTGATAAAGAAATTGGTTCTAATAAGGTTGAGGA TAAGTGTGGTGTCTGTGGAGGAGATAATTCCCACTGCCGAACCGTGAAGGGGACATTT ACCAGAACTCCCAGGAAGCTTGGGTACCTTAAGATGTTTGATATACCCCCTGGGGCTA GACATGTGTTAATCCAAGAAGACGAGGCTTCTCCTCATATTCTTGCTATTAAGAACCA GGCTACAGGCCATTATATTTTAAATGGCAAAGGGGAGGAAGCCAAGTCGCGGACCTTC ATAGATCTTGGTGTGGAGTGGGATTATAACATTGAAGATGACATTGAAAGTCTTCACA CCGATGGACCTTTACATGATCCTGTTATTGTTTTGATTATACCTCAAGAAAATGATAC CCGCTCTAGCCTGACATATAAGTACATCATCCATGAAGACTCTGTACCTACAATCAAC AGCAACAATGTCATCCAGGAAGAATTAGATACTTTTGAGTGGGCTTTGAAGAGCTGGT CTCAGTGTTCCAAACCCTGTGGTGGAGGTTTCCAGTACACTAAATATGGATGCCGTAG GAAAAGTGATAATAAAATGGTCCATCGCAGCTTCTGTGAGGCCAACAAAAAGCCGAAA CCTATTAGACGAATGTGCAATATTCAAGAGTGTACACATCCACTCTGGGTAGCAGAAG AATGGGAACACTGCACCAAAACCTGTGGAAGTTCTGGCTATCAGCTTCGCACTGTACG CTGCCTTCAGCCACTCCTTGATGGCACCAACCGCTCTGTGCACAGCAAATACTGCATG GGTGACCGTCCCGAGAGCCGCCGGCCCTGTAACAGAGTGCCCTGCCCTGCACAGTGGA AAACAGGACCCTGGAGTGAGTGTTCAGTGACCTGCGGTGAAGGAACGGAGGTGAGGCA GGTCCTCTGCAGGGCTGGGGACCACTGTGATGGTGAAAAGCCTGAGTCGGTCAGAGCC TGTCAACTGCCTCCTTGTCTCGAG

ORF Start: at 1		ORF Stop: end of sequence	
SEQ ID NO: 28		762 aa	MW at 86650.1kD
NOV8c, 210387874 Protein Sequence	GSIEVLLGVDDSVVRFHGKEHVQNYLLTLMNIVNEIYHDESLGVHINVVLVRMIMLGY AKSISLIERGNPSRSLENVCRWASQQQRSDLNHSEHHDHAIFLTRQDFGPAGMQGYAP VTGMCHPVRSCTLNHEDGFSSAFVVAHETGHVLGMEHDGQGNRCGDETAMGSVMAPLV QAAFHRYHWSRCSGQELKRYIHSYDCLLDDPFDHDWPKLPELPGINYSMDEQCRFDFG VGYKMCTAFRTFDPCKQLWCSHPDNPYFCKTKKGPPLDGTECAAGKWCYKGHCMWKNA NQQKQDGNWGSWTKFGSCSRTCGTGVRFRTRQCNNPMPINGGQDCPGVNFEYQLCNTE ECQKHFEDFRAQQCQQRNSHFEYQNTKHHWLPYEHPDPKKRCHLYCQSKETGDVAYMK QLVHDGTHCSYKDPYSICVRGECVKVGCDKEIGSNKVEDKCGVCGGDNSHCRTVKGTF TRTPRKLGYLKMFDIPPGARHVLIQEDEASPHILAIKNQATGHYILNGKGEEAKSRTF IDLGVEWDYNIEDDIESLHTDGPLHDPVIVLIIPQENDTRSSLTYKYIIHEDSVPTIN SNNVIQEELDTFEWALKSWSQCSKPCGGGFQYTKYGCRRKSDNKMVHRSFCEANKKPK PIRRMCNIQECTHPLWVAEEWEHCTKTCGSSGYQLRTVRCLQPLLDGTNRSVHSKYCM GDRPESRRPCNRVPCPAQWKTGPWSECSVTCGEGTEVRQVLCRAGDHCDGEKPESVRA CQLPPCLE		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 8B.

Table 8B. Comparison of NOV8a against NOV8b and NOV8c.		
Protein Sequence	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV8b	258..1015 1..758	749/758 (98%) 749/758 (98%)
NOV8c	257..1015 2..760	758/759 (99%) 759/759 (99%)

Further analysis of the NOV8a protein yielded the following properties shown in Table 8C.

Table 8C. Protein Sequence Properties NOV8a	
PSort analysis:	0.5708 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 21 and 22

A search of the NOV8a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 8D.

Table 8D. Geneseq Results for NOV8a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV8a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB73550	Human ADAM-type metalloprotease MDTS5, SEQ ID NO:10 - Homo sapiens, 1205 aa. [JP2001017183-A, 23-JAN-2001]	1..1205 1..1205	1204/1205 (99%) 1205/1205 (99%)	0.0
AAB21254	Human metalloproteinase KIAA0366 - Homo sapiens, 1201 aa. [WO200053774-A2, 14-SEP-2000]	5..1205 1..1201	1199/1201 (99%) 1200/1201 (99%)	0.0
AAU72895	Human metalloprotease partial protein sequence #7 - Homo sapiens, 1186 aa. [WO200183782-A2, 08-NOV-2001]	41..1128 37..1139	650/1129 (57%) 792/1129 (69%)	0.0
AAW47028	Human N-proteinase (130 kDa long form) - Homo sapiens, 1211 aa. [WO9800555-A1, 08-JAN-1998]	35..1051 49..1094	655/1070 (61%) 796/1070 (74%)	0.0
AAU74750	Human protease PRTS-10 protein sequence - Homo sapiens, 1189 aa. [WO200198468-A2, 27-DEC-2001]	41..1128 37..1142	651/1131 (57%) 793/1131 (69%)	0.0

In a BLAST search of public sequence databases, the NOV8a protein was found to have homology to the proteins shown in the BLASTP data in Table 8E.

	Table 8E. Public BLASTP Results for NOV8a			
Protein Accession Number	Protein/Organism/Length	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O15072	ADAM-TS 3 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 3) (ADAMTS-3) (ADAM-TS3) - Homo sapiens (Human), 1201 aa (fragment).	5..1205 1..1201	1199/1201 (99%) 1200/1201 (99%)	0.0
O95450	ADAM-TS 2 precursor (EC 3.4.24.14) (A disintegrin and metalloproteinase with thrombospondin motifs 2) (ADAMTS-2) (ADAM-TS2) (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (pNPI) (Procollagen I/II amino-propeptide processing enzyme) - Homo sapiens (Human), 1211 aa.	35..1051 49..1094	654/1070 (61%) 795/1070 (74%)	0.0
P79331	ADAM-TS 2 precursor (EC 3.4.24.14) (A disintegrin and metalloproteinase with thrombospondin motifs 2) (ADAMTS-2) (ADAM-TS2) (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (pNPI) - Bos taurus (Bovine), 1205 aa.	44..1091 50..1120	655/1101 (59%) 803/1101 (72%)	0.0
AAL79814	ADAMTS14 - Homo sapiens (Human), 1159 aa.	80..1128 39..1112	641/1097 (58%) 778/1097 (70%)	0.0
Q8WXS8	A DISINTEGRIN-LIKE AND METALLOPROTEASE WITH THROMBOSPONDIN TYPE 1 MOTIF 14 PRECURSOR - Homo sapiens (Human), 1223 aa.	41..1128 37..1176	650/1166 (55%) 793/1166 (67%)	0.0

PFam analysis predicts that the NOV8a protein contains the domains shown in the Table 8F.

Table 8F. Domain Analysis of NOV8a			
Pfam Domain	**NOV8a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Pep_M12B_propep	94..226	31/147 (21%) 97/147 (66%)	6.3e-08
Reprolysin	258..460	60/218 (28%) 143/218 (66%)	1.2e-08
tsp_1	555..605	21/54 (39%) 34/54 (63%)	4.1e-10
tsp_1	847..904	13/63 (21%) 43/63 (68%)	0.00094
tsp_1	909..966	17/62 (27%) 39/62 (63%)	0.033
tsp_1	970..1015	19/54 (35%) 35/54 (65%)	2.2e-08

EXAMPLE 9.

The NOV9 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 9A.

Table 9A. NOV9 Sequence Analysis	
SEQ ID NO: 29	937 bp
NOV9a, CG95145-01 DNA Sequence	GGCAGCGGCGGCGGCGTGCCTGCCTGGCGGCCGTCGGCGTACTCTTGGCCATGGCGCT CGGGCTGCTCATCGCCGTGCCGCTGCTGCTGCAGGCGGCGCCCCGAGGCGCCGCGCAC TATGAGATGATGGGCACCTGCCGCATGATCTGCGACCCTTACACTGCCGCGCCCGGCG GGGAGCCCCCGGGTGCAAAGGCGCAGCCACCCGGACCCAGCACCGCCGCCCTGGAAGT CATGCAGGACCTCAGCGCCAACCCTCCTCCTCCTTTCATCCAGGGACCCAAGGGCGAC CCGGGGCGACCGGGCAAGCCAGGGCCGCGGGGGCCCCCTGGAGAGCCGGGCCCGCCTG GACCCAGGGGCCCTCCGGGAGAGAAGGGCGACTCGGGGCGGCCCGGGCTGCCAGGGCT GCAACTGACGGCGGGCACGGCCAGCGGCGTCGGGGTGGTGGGCGGCGGGGCCGGGGTA GGTGGCGATTCCGAGGGTGAAGTGACCAGTGCGCTGAGCGCCACCTTCAGCGGCCCCA AGATCGCCTTCTATGTGGGTCTCAAGAGCCCCACGAAGGCTATGAGGTGCTGAAGTT CGATGACGTGGTCACCAACCTCGGCAATCACTATGACCCCACCACGGGCAAGTTCAGC TGCCAGGTACGCGGCATCTACTTCTTCACCTACCACATCCTCATGCGCGGCGGCGACG GCACCAGCATGTGGGCGGACCTCTGCAAGAACGGGCAGGTCCGGGCCAGCGCCATTGC ACAGGACGCCGACCAGAACTACGACTACGCCAGTAACAGCGTGGTGCTGCACTTGGAT TCAGGGGACGAAGTGTATGTGAAGCTGGATGGCGGGAAGGCTCACGGAGGCAATAATA ACAAGTACAGCACGTTCTCGGGCTTTCTTCTGTACCCGGATTAGGGGCGCGGGGGGTG CGAGGCGGG
	ORF Start: ATG at 51	ORF Stop: TAG at 912
	SEQ ID NO: 30	287 aa	MW at 29467.8kD
NOV9a, CG95145-01 Protein Sequence	MALGLLIAVPLLLQAAPRGAAHYEMMGTCRMICDPYTAAPGGEPPGAKAQPPGPSTAA LEVMQDLSANPPPPFIQGPKGDPGRPGKPGPRGPPGEPGPPGPRGPPGEKGDSGRPGL PGLQLTAGTASGVGVVGGGAGVGGDSEGEVTSALSATFSGPKIAFYVGLKSPHEGYEV LKFDDVVTNLGNHYDPTTGKFSCQVRGIYFFTYHILMRGGDGTSMWADLCKNGQVRAS AIAQDADQNYDYASNSVVLHLDSGDEVYVKLDGGKAHGGNNNKYSTFSGFLLYPD

Further analysis of the NOV9a protein yielded the following properties shown in Table 9B.

Table 9B. Protein Sequence Properties NOV9a	
PSort analysis:	0.3798 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Cleavage site between residues 22 and 23

A search of the NOV9a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 9C.

Table 9C. Geneseq Results for NOV9a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU09865	Novel human secreted protein #6 – Homo sapiens, 287 aa. [WO200179454-A1, 25-OCT-2001]	1..287 1..287	270/287 (94%) 273/287 (95%)	e-164
ABB53290	Human polypeptide #30 – Homo sapiens, 255 aa. [WO200181363-A1, 01-NOV-2001]	1..287 1..255	194/290 (66%) 219/290 (74%)	e-105
AAG64212	Murine HSP47 interacting protein, #2 – Mus sp, 255 aa. [JP2001145493-A, 29-MAY-2001]	1..287 1..255	193/290 (66%) 218/290 (74%)	e-104
AAM40607	Human polypeptide SEQ ID NO 5538 – Homo sapiens, 255 aa. [WO200153312-A1, 26-JUL-2001]	65..287 31..252	84/233 (36%) 115/233 (49%)	1e-29
AAM38821	Human polypeptide SEQ ID NO 1966 – Homo sapiens, 253 aa. [WO200153312-A1, 26-JUL-2001]	65..287 29..250	84/233 (36%) 115/233 (49%)	1e-29

In a BLAST search of public sequence databases, the NOV9a protein was found to have homology to the proteins shown in the BLASTP data in Table 9D.

Table 9D. Public BLASTP Results for NOV9a				
Protein Accession Number	Protein/Organism/Length	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9ESN4	Gliacolin precursor - Mus musculus (Mouse), 255 aa.	1..287 1..255	193/290 (66%) 218/290 (74%)	e-104
O75973	C1q-related factor precursor - Homo sapiens (Human), 258 aa.	1..287 1..258	179/288 (62%) 213/288 (73%)	e-101
O88992	C1q-related factor precursor - Mus musculus (Mouse), 258 aa.	1..287 1..258	177/288 (61%) 210/288 (72%)	e-100
AAH22724	HYPOTHETICAL 13.1 KDA PROTEIN - Mus musculus (Mouse), 120 aa (fragment).	168..287 1..120	102/120 (85%) 114/120 (95%)	6e-60
S31216	collagen alpha 1(X) chain precursor - mouse, 680 aa.	38..286 439..679	94/260 (36%) 129/260 (49%)	3e-32

PFam analysis predicts that the NOV9a protein contains the domains shown in the Table 9E.

Table 9E. Domain Analysis of NOV9a			
Pfam Domain	NOV9a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Collagen	76..135	31/60 (52%) 44/60 (73%)	0.00022
C1q	160..284	47/140 (34%) 93/140 (66%)	7.6e-31

5

EXAMPLE 10.

The NOV10 clone was analyzed, and the nucleotide and encoded polypeptide
sequences are shown in Table 10A.

Table 10A. NOV10 Sequence Analysis		
	SEQ ID NO: 31	3121 bp
NOV10a, CG95250-01 DNA Sequence	TTTTGACAGCTGCCACAGTCTCTGAGCTCCAGCCTCGCGCCTGAACCCGGTCCCTGCC ATGGGGCCCCCTTCCAGCTCAGGCTTCTATGTGAGCCGCGCAGTGGCCCTGCTGCTGG CTGGGCTGGTAGCCGCCCTCCTGCTGGCGCTGGCCGTACTCGCCGCCTTGTACGGCCA CTGCGAGCGCGTCCCACCGTCGGAGCTGCCTGGACTCAGGGACTTGGAAGCCGAGTCT TCCCCTCCCCTCAGGCAGAAGCCGACGCCAACCCCGAAACCCAGCAGTGCACGCGAGC TAGCGGTGACGACCACCCCGAGCAACTGGCGACCCCCGGGGCCCTGGGACCAGCTACG CCTGCCGCCCTGGCTCGTGCCGCTGCACTACGATCTGGAGCTGTGGCCGCAGCTGAGG CCCGACGAGCTTCCGGCCGGGTCTTTGCCCTTCACTGGCCGCGTGAACATCACGGTGC GCTGCACGGTGGCCACCTCTCGACTGCTGCTGCATAGCCTCTTCCAGGACTGCGAGCG CGCCGAGGTGCGGGGACCCCTTTCCCCGGGCACTGGGAACGCCACAGTGGGCCGCGTG CCCGTGGACGACGTGTGGTTCGCGCTGGACACGGAATACATGGTGCTGGAGCTCAGTG AGCCCCTGAAACCTGGTAGCAGCTACGAGCTGCAGCTTAGCTTCTCGGGCCTGGTGAA GGAAGACCTCAGGGAGGGACTCTTCCTCAACGTCTACACCGACCAGGGCGAGCGCAGG GCCCTGTTAGCGTCCCAGCTGGAACCAACATTTGCCAGGTATGTTTTCCCTTGTTTTG ATGAGCCAGCTCTGAAGGCAACTTTTAATATTACAATGATTCATCATCCAAGTTATGT GGCCCTTTCCAACATGCCAAAGAATTCTCAGTCTGAAAAAGAAGATGTGAATGGAAGC AAATGGACTGTTACAACCTTTTCCACTACGCCCCACATGCCAACTTACTTAGTCGCAT TTGTTATATGTGACTATGACCACGTCAACAGAACAGAAAGGGGCAAGGAGGTGATACG CATCTGGGCCCGGAAAGATGCAATTGCAAATGGAAGTGCAGACTTTGCTTTGAACATC ACAGGTCCCATCTTCTCTTTTCTGGAGGATTTGTTTAATATCAGTTACTCTCTTCCAA AAACAGATATAATTGCCTTGCCTAGTTTTGACAACCATGCAATGGAAAACTGGGGACT AATGATATTTGATGAATCAGGATTGTTGTTGGAACCAAAAGATCAACTGACAGAAAAA AAGACTCTGATCTCCTATGTTGTCTCCCACGAGATTGGACACCAGTGGTTTGGAAACT TGGTTACCATGAATTGGTGGAACAATATCTGGCTCAACGAGGGTTTTGCATCTTATTT TGAGTTTGAAGTAATTAACTACTTTAATCCTAAACTCCCAAGAGTAAGTAATGAGATC TTTTTTTCTAACATTTTACATAATATCCTCAGAGAAGATCACGCCCTGGTGACTAGAG CTGTGGCCATGAAGGTGGAAAATTTCAAAACAAGTGAAATACAGGAACTCTTTGACAT ATTTACTTACAGCAAGGTAAAAGCAGTTAGAAATTTCCTTTGGTTTTGTACTCTGGTA GAAAGTCTATATCATCATACATTACAGTCATATTTGAAGACATTTTCCTACTCAAACG CTGAGCAAGATGATCTATGGAGGCATTTTCAACAGGCCATAGATGACCAGAGTACAGT TATTTTGCCAGCAACAATAAAAAACATAATGGACAGTTGGACACACCAGAGTGGTTTT CCAGTGATCACTTTAAATGTGTCTACTGGCGTCATGAAACAGGAGCCATTTTATCTTG AAAACATTAAAAATCGGACTCTTCTAACCAGCAAGGACACATGGATTGTCCCTATTCT TTGGATAAAAAATGGAACTACACAACCTTTAGTCTGGCTAGATCAAAGCAGCAAAGTA TTCCCAGAAATGCAAGTTTCAGATTCTGACCATGACTGGGTGATTTTGAATTTGAATA TGACTGGATATTATAGAGTTAATTATGATAAATTAGGTTGGAAGAAACTAAATCAACA ACTTGAAAAGGATCCTAAGGCTATTCCTGTTATTCACAGACTGCAGTTGATTGATGAT GCCTTTTCCTTGTCTAAGAAGTTATTGAGCTTGTCCCGAACTTTGCCTTTGGACCACT TCTTCTTTTTGGCCTTGCCCCCGGATTTGTTCACTGGGTCTTTGTCTTTCTTGGCTGA CTTTCCAGCGTCCTTCTTCTCGTCGTCCTTGGGCGCTGCTCAGGAAAAAAAAGATTC CTTCCAAAATATTACTGCTCATTGATGATGCACTTGGTTACACAAGCACTAATGATGG AGATGTACAAGGAGATTAATGTTGTTTTCATGCCCACTAACACAATATCAATTCTTCA GCCCAGGGATCAAGGAGTAACTGCGTGTTGGTTGGGCCTTGAAGACTGCCTTCAGCTG TCAAAAGAACTTTTCGCAAAATGGGTGGATCATCCAGAAAATAGAATACCTTATCCAA TTAAAGATGTGGTTTTATGTTATGGCATTGCCTTGGGAAGTGATAAAGAGTGGGACAT CTTGTTAAATACTTACACTAATACAACAAACAAGAAGAAAGATTCAACTTGCTTAT GCAATGAGCTGCAGCAAAGACCCATGGATACTTAACAGGAGATATATGGAGTATGCCA TCAGCACATCTCCATTCACTTCTAATGAAACAAATATAATTGAGGTTGTGGCTTCATC TGAAGTTGGCCGGTATGTCGCAAAGACTTCTTAGTCAACAACTGGCAAGCTGTGAGT CACACTTTGAGCAGGTATGGAACACAATCATTGATTAATCTAATATATACAATAGGGA

	GAACCGTAACTACAGATTTACAGATTGTGGAGCTGCAGCAGTTTTTTCAGTAACATGTT GGAGGAACACCAGAGGATCAGAGTTCATGCCAACTTACAGACAATAAAGAATGAAAAT CTGAAAAACAAGAAGCTAAGTGCCAGGATAGCTGCGTGGCTAAGGAGAAACAC**ATAGC** TTGTGGCTATCTTTCAGCACTCCTCTTGCATATTATAATGTAGTTTG
	ORF Start: ATG at 59 ORF Stop: TAG at 3071
	SEQ ID NO: 32 1004 aa MW at 114692.2kD
NOV10a, CG95250-01 Protein Sequence	MGPPSSSGFYVSRAVALLLAGLVAALLLALAVLAALYGHCERVPPSELPGLRDLEAES SPPLRQKPTPTPKPSSARELAVTTTPSNWRPPGPWDQLRLPPWLVPLHYDLELWPQLR PDELPAGSLPFTGRVNITVRCTVATSRLLLHSLFQDCERAEVRGPLSPGTGNATVGRV PVDDVWFALDTEYMVLELSEPLKPGSSYELQLSFSGLVKEDLREGLFLNVYTDQGERR ALLASQLEPTFARYVFPCFDEPALKATFNITMIHHPSYVALSNMPKNSQSEKEDVNGS KWTVTTFSTTPHMPTYLVAFVICDYDHVNRTERGKEVIRIWARKDAIANGSADFALNI TGPIFSFLEDLFNISYSLPKTDIIALPSFDNHAMENWGLMIFDESGLLLEPKDQLTEK KTLISYVVSHEIGHQWFGNLVTMNWWNNIWLNEGFASYFEFEVINYFNPKLPRVSNEI FFSNILHNILREDHALVTRAVAMKVENFKTSEIQELFDIFTYSKVKAVRNFLWFCTLV ESLYHHTLQSYLKTFSYSNAEQDDLWRHFQQAIDDQSTVILPATIKNIMDSWTHQSGF PVITLNVSTGVMKQEPFYLENIKNRTLLTSKDTWIVPILWIKNGTTQPLVWLDQSSKV FPEMQVSDSDHDWVILNLNMTGYYRVNYDKLGWKKLNQQLEKDPKAIPVIHRLQLIDD AFSLSKKLLSLSRTLPLDHFFFLALPPDLFTGSLSFLADFPASFFSSSLGAAQEKKRF LPKYYCSLMMHLVTQALMMEMYKEINVVFMPTNTISILQPRDQGVTACWLGLEDCLQL SKELFAKWVDHPENRIPYPIKDVVLCYGIALGSDKEWDILLNTYTNTTNKEEKIQLAY AMSCSKDPWILNRRYMEYAISTSPFTSNETNIIEVVASSEVGRYVAKDFLVNNWQAVS HTLSRYGTQSLINLIYTIGRTVTTDLQIVELQQFFSNMLEEHQRIRVHANLQTIKNEN LKNKKLSARIAAWLRRNT
	SEQ ID NO: 33 2880 bp
NOV10b, CG95250-02 DNA Sequence	TTTTGACAGCTGCCACAGTCTCTGAGCTCCAGCCTCGCGCCTGAACCCGGTCCCTGCC **ATG**GGGCCCCCTTCCAGCTCAGGCTTCTATGTGAGCCGCGCAGTGGCCCTGCTGCTGG CTGGGCTGGTAGCCGCCCTCCTGCTGGCGCTGGCCGTACTCGCCGCCTTGTACGGCCA CTGCGAGCGCGTCCCACCGTCGGAGCTGCCTGGACTCAGGGACTTGGAAGCCGAGTCT TCCCCTCCCCTCAGGCAGAAGCCGACGCCAACCCCGAAACCCAGCAGTGCACGCGAGC TAGCGGTGACGACCACCCCGAGCAACTGGCGACCCCCGGGGCCCTGGGACCAGCTACG CCTGCCGCCCTGGCTCGTGCCGCTGCACTACGATCTGGAGCTGTGGCCGCAGCTGAGG CCCGACGAGCTTCCGGCCGGGTCTTTGCCCTTCACTGGCCGCGTGAACATCACGGTGC GCTGCACGGTGGCCACCTCTCGACTGCTGCTGCATAGCCTCTTCCAGGACTGCGAGCG CGCCGAGGTGCGGGGACCCCTTTCCCCGGGCACTGGGAACGCCACAGTGGGCCGCGTG CCCGTGGACGACGTGTGGTTCGCGCTGGACACGGAATACATGGTGCTGGAGCTCAGTG AGCCCCTGAAACCTGGTAGCAGCTACGAGCTGCAGCTTAGCTTCTCGGGCCTGGTGAA GGAAGACCTCAGGGAGGGACTCTTCCTCAACGTCTACACCGACCAGGGCGAGCGCAGG GCCCTGTTAGCGTCCCAGCTGGAACCAACATTTGCCAGGTATGTTTTCCCTTGTTTTG ATGAGCCAGCTCTGAAGGCAACTTTTAATATTACAATGATTCATCATCCAAGTTATGT GGCCCTTTCCAACATGCCAAAGAATTCTCAGTCTGAAAAAGAAGATGTGAATGGAAGC AAATGGACTGTTACAACCTTTTCCACTACGCCCCACATGCCAACTTACTTAGTCGCAT TTGTTATATGTGACTATGACCACGTCAACAGAACAGAAAGGGGCAAGGAGGTGATACG CATCTGGGCCCGGAAAGATGCAATTGCAAATGGAAGTGCAGACTTTGCTTTGAACATC ACAGGTCCCATCTTCTCTTTTCTGGAGGATTTGTTTAATATCAGTTACTCTCTTCCAA AAACAGATATAATTGCCTTGCCTAGTTTTGACAACCATGCAATGGAAAACTGGGGACT AATGATATTTGATGAATCAGGATTGTTGTTGGAACCAAAAGATCAACTGACAGAAAAA AAGACTCTGATCTCCTATGTTGTCTCCCACGAGATTGGACACCAGTGGTTTGGAAACT TGGTTACCATGAATTGGTGGAACAATATCTGGCTCAACGAGGGTTTTGCATCTTATTT TGAGTTTGAAGTAATTAACTACTTTAATCCTAAACTCCCAAGAGTAAGTAATGAGATC TTTTTTTCTAACATTTTACATAATATCCTCAGAGAAGATCACGCCCTGGTGACTAGAG CTGTGGCCATGAAGGTGGAAAATTTCAAAACAAGTGAAATACAGGAACTCTTTGACAT ATTTACTTACAGCAAGGGAGCGTCTATGGCCCGGATGCTTTCTTGTTTCTTGAATGAG CATTTATTTGTCAGTGCATTACAGTCATATTTGAAGACATTTTCCTACTCAAACGCTG AGCAAGATGATCTATGGAGGCATGATTTTTTAAAACAGGCCATAGATGACCAGAGTAC AGTTATTTTGCCAGCAACAATAAAAAACATAATGGACAGTTGGACACACCAGAGTGGT TTTCCAGTGATCACTTTAAATGTGTCTACTGGCGTCCTGATACAGGAGCCATTTTATC

	TTGAAAACATTAAAAATCGGACTCTTCTAACCAGCAATGACACATGGATTGTCCCTAT TCTTTGGATAAAAAATGGAACTACACAACCTTTAGTCTGGCTAGATCAAAGCAGCAAA GTATTCCCAGAAATGCAAGTTTCAGATTCTGACCATGACTGGGTGATTTTGAATTTGA ATATGACTGGATATTATAGAGTTAATTATGATAAATTAGGTTGGAAGAAACTAAATCA ACAACTTGAAAAGGATCCTAAGGCTATTCCTGTTATTCACAGACTGCAGTTCATTGAT GATGCCTTTTCCTTGTCTAAAAACAATTATATTGAGATTGAAACAGCACTTGAGTTAA CCAAGTACCTTGCTGAAGAAGATGAAATTATAGTATGGCATACAGTCTTGGTAAACTT GGTAACCAGGGATCTTGTTTCTGAGGTAACATCTATGATATATACTCATTATTAAAG ACTGCGTGTTGGTTGGGCCTTGAAGACTGCCTTCAGCTGTCAAAAGAACTTTTCGCAA AATGGGTGGATCATCCAGAAAATATTAAAGATGTGGTTTTATGTTATGGCATTGCCTT GGGAAGTGATAAAGAGTGGGACATCTTGTTAAATACTTACACTAATACAACAAACAAA GAAGAAAAGATTCAACTTGCTTATGCAATGAGCTGCAGCAAAGACCCATGGATACTTA ACAGGTATATGGAGTATGCCATCAGCACATCTCCATTCACTTCTAATGAAACAAATAT AATTGAGGTTGTGGCTTCATCTGAAGTTGGCCGGTATGTCGCAAAGACTTCTTAGTC AACAACTGGCAAGCTGTGAGTAAAAGG<u>TAAGAAGGAAAGTGAGACCTTTCTTTCATTT AGGCCACTGGTTTGGCACTGGAAGCTCAGCTTTAGTCTAGCTTGGAAGCTCAGCTTTA GTCTAGCTAGGCCACAAACGTCCTTTGCATTGACTAGAAAGTTATCATTTTTCCTTT GTTTAGTCTCACTACAAACTGCCTGTGTTATGGAAGAG</u>
	ORF Start: ATG at 59 **ORF Stop: TAA at 2696**
	SEQ ID NO: 34 **879 aa** **MW at 100068.0kD**
NOV10b, CG95250-02 Protein Sequence	MGPPSSSGFYVSRAVALLLAGLVAALLLALAVLAALYGHCERVPPSELPGLRDLEAES SPPLRQKPTPTPKPSSARELAVTTTPSNWRPPGPWDQLRLPPWLVPLHYDLELWPQLR PDELPAGSLPFTGRVNITVRCTVATSRLLLHSLFQDCERAEVRGPLSPGTGNATVGRV PVDDVWFALDTEYMVLELSEPLKPGSSYELQLSFSGLVKEDLREGLFLNVYTDQGERR ALLASQLEPTFARYVFPCFDEPALKATFNITMIHHPSYVALSNMPKNSQSEKEDVNGS KWTVTTFSTTPHMPTYLVAFVICDYDHVNRTERGKEVIRIWARKDAIANGSADFALNI TGPIFSFLEDLFNISYSLPKTDIIALPSFDNHAMENWGLMIFDESGLLLEPKDQLTEK KTLISYVVSHEIGHQWFGNLVTMNWWNNIWLNEGFASYFEFEVINYFNPKLPRVSNEI FFSNILHNILREDHALVTRAVAMKVENFKTSEIQELFDIFTYSKGASMARMLSCFLNE HLFVSALQSYLKTFSYSNAEQDDLWRHDFLKQAIDDQSTVILPATIKNIMDSWTHQSG FPVITLNVSTGVLIQEPFYLENIKNRTLLTSNDTWIVPILWIKNGTTQPLVWLDQSSK VFPEMQVSDSDHDWVILNLNMTGYYRVNYDKLGWKKLNQQLEKDPKAIPVIHRLQFID DAFSLSKNNYIEIETALELTKYLAEEDEIIVWHTVLVNLVTRDLVSEVNIYDIYSLLK TACWLGLEDCLQLSKELFAKWVDHPENIKDVVLCYGIALGSDKEWDILLNTYTNTTNK EEKIQLAYAMSCSKDPWILNRYMEYAISTSPFTSNETNIIEVVASSEVGRYVAKDFLV NNWQAVSKR
	SEQ ID NO: 35 **1695 bp**
NOV10c, CG95250-03 DNA Sequence	ATGGGGCCCCCTTCCAGCTCAGGCTTCTATGTGAGCCGCGCAGTGGCCCTGCTGCTGG CTCGGCTGGTAGCCGCCCTCCTGCTGGCGCTGGCCGTACTCGCCGCCTTGTACGGCCA CTGCGAGCGCGTCCCACCGTCGGAGCTGCCTGGACTCAGGGACTTGGAAGCCGAGTCT TCCCCTCCCCTCAGGCAGAAGCCGACGCCAACCCCGAAACCCAGCAGTGCACGCGAGC TAGCGGTGACGACCACCCCGAGCAACTGGCGACCCCCGGGGCCCTGGGACCAGCTACG CCTGCCGCCCTGGCTCGTGCCGCTGCACTACGATCTGGAGCTGTGGCCGCAGCTGAGG CCCGACGAGCTTCCGGCCGGGTCTTTGCCCTTCACTGGCCGCGTGAACATCACGGTGC GCTGCACGGTGGCCACCTCTCGACTGCTGCTGCATAGCCTCTTCCAGGACTGCGAGCG CGCCGAGGTGCGGGGACCCCTTTCCCCGGGCACTGGGAACGCCACAGTGGGCCGCGTG CCCGTGGACGACGTGTGGTTCGCGCTGGACACGGAATACATGGTGCTGGAGCTCAGTG AGCCCCTGAAACCTGGTAGCAGCTACGAGCTGCAGCTTAGCTTCTCGGGCCTGGTGAA GGAAGACCTCAGGGAGGGACTCTTCCTCAACGTCTACACCGACCAGGGCGAGCGCAGG GCCCTGTTAGCGTCCCAGCTGGAACCAACATTTGCCAGGTATGTTTTCCCTTGTTTTG ATGAGCCAGCTCTGAAGGCAACTTTTAATATTACAATGATTCATCATCCAAGTTATGT GGCCCTTTCCAACATGCCAAAGCTAGGTCAGTCTGAAAAGAAGATGTGAATGGAAGC AAATGGACTGTTACAACCTTTTCCACTACGCCCCACATGCCAACTTACTTAGTCGCAT TTGTTATATGTGACTATGACCACGTCAACAGAACAGAAAGGGGCAAGGAGATACGCAT CTGGGCCCGGAAAGATGCAATTGCAAATGGAAGTGCAGACTTTGCTTTGAACATCACA GGTCCCATCTTCTCTTTTCTGGAGGATTTGTTTAATATCAGTTACTCTCTTCCAAAAA CAGATATAATTGCCTTGCCTAGTTTTGACAACCATGCAATGGAAAACTGGGGACTAAT

GATATTTGATGAATCAGGATTGTTGTTGGAACCAAAAGATCAACTGACAGAAAAAAG ACTCTGATCTCCTATGTTGTCTCCCACGAGATTGGACACCAGTGGTTTGGAAACTTGG TTACCATGAATTGGTGGAACAATATCTGGCTCAACGAGGGTTTTGCATCTTATTTTGA GTTTGAAGTAATTAACTACTTTAATCCTAAACTCCCAAGAGTAAGTATCTTTTTTTCT AACATTTTACATAATATCCTCAGAGAAGATCACGCCCTGGTGACTAGAGCTGTGGCCA TGAAGGTGGAAAATTTCAAAACAAGTGAAATACAGGAACTCTTTGACATATTTACTTA CAGCAAGGGAGCGTCTATGGCCCGGATGCTTTCTTGTTTCTTGAATGAGCATTTATTT GTCAGTGCATTACAGTCATATTTGAAGACATTTTCCTACTCAAACGCTGAGCAAGATG ATCTATGGAGGCATTTTCAAATGGTAATTGTCCTACTTTCTGACACATTCTTGCTGAG TTGTTTTGTA**TGA**
ORF Start: ATG at 1	ORF Stop: TGA at 1693
SEQ ID NO: 36	564 aa	MW at 63684.2kD
NOV10c, CG95250-03 Protein Sequence	MGPPSSSGFYVSRAVALLLAGLVAALLLALAVLAALYGHCERVPPSELPGLRDLEAES SPPLRQKPTPTPKPSSARELAVTTTPSNWRPPGPWDQLRLPPWLVPLHYDLELWPQLR PDELPAGSLPFTGRVNITVRCTVATSRLLLHSLFQDCERAEVRGPLSPGTGNATVGRV PVDDVWFALDTEYMVLELSEPLKPGSSYELQLSFSGLVKEDLREGLFLNVYTDQGERR ALLASQLEPTFARYVFPCFDEPALKATFNITMIHHPSYVALSNMPKLGQSEKEDVNGS KWTVTTFSTTPHMPTYLVAFVICDYDHVNRTERGKEIRIWARKDAIANGSADFALNIT GPIFSFLEDLFNISYSLPKTDIIALPSFDNHAMENWGLMIFDESGLLLEPKDQLTEKK TLISYVVSHEIGHQWFGNLVTMNWWNNIWLNEGFASYFEFEVINYFNPKLPRVSIFFS NILHNILREDHALVTRAVAMKVENFKTSEIQELFDIFTYSKGASMARMLSCFLNEHLF VSALQSYLKTFSYSNAEQDDLWRHFQMVIVLLSDTFLLSCFV

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 10B.

Table 10B. Comparison of NOV10a against NOV10b and NOV10c.

Protein Sequence	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV10b	1..928 1..877	783/931 (84%) 808/931 (86%)
NOV10c	1..555 1..551	509/555 (91%) 512/555 (91%)

Further analysis of the NOV10a protein yielded the following properties shown in Table 10C.

Table 10C. Protein Sequence Properties NOV10a

PSort analysis:	0.8000 probability located in mitochondrial inner membrane; 0.6500 probability located in plasma membrane; 0.6199 probability located in microbody (peroxisome); 0.3000 probability located in Golgi body
SignalP analysis:	Cleavage site between residues 35 and 36

5   A search of the NOV10a protein against the Geneseq database, a proprietary

database that contains sequences published in patents and patent publication, yielded

several homologous proteins shown in Table 10D.

Table 10D. Geneseq Results for NOV10a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU72905	Human metalloprotease partial protein sequence #17 - Homo sapiens, 990 aa. [WO200183782-A2, 08-NOV-2001]	1..1004 1..990	890/1009 (88%) 915/1009 (90%)	0.0
AAW93621	Human CD13/aminopeptidase N protein - Homo sapiens, 967 aa. [WO9913329-A1, 18-MAR-1999]	6..962 2..921	337/976 (34%) 517/976 (52%)	e-152
AAB54345	Human pancreatic cancer antigen protein sequence SEQ ID NO:797 - Homo sapiens, 977 aa. [WO200055320-A1, 21-SEP-2000]	6..962 12..931	336/976 (34%) 517/976 (52%)	e-151
ABG20442	Novel human diagnostic protein #20433 - Homo sapiens, 935 aa. [WO200175067-A2, 11-OCT-2001]	6..961 2..889	329/976 (33%) 504/976 (50%)	e-139
ABG20442	Novel human diagnostic protein #20433 - Homo sapiens, 935 aa. [WO200175067-A2, 11-OCT-2001]	6..961 2..889	329/976 (33%) 504/976 (50%)	e-139

In a BLAST search of public sequence databases, the NOV10a protein was found to have homology to the proteins shown in the BLASTP data in Table 10E.

Table 10E. Public BLASTP Results for NOV10a				
Protein Accession Number	Protein/Organism/Length	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9D633	4833403I15RIK PROTEIN - Mus musculus (Mouse), 559 aa.	1..555 1..549	362/557 (64%) 423/557 (74%)	0.0
P15541	Aminopeptidase N (EC 3.4.11.2) (Microsomal aminopeptidase) (Leukemia antigen CD13) - Oryctolagus cuniculus (Rabbit), 965 aa.	6..968 1..925	347/986 (35%) 520/986 (52%)	e-162
A32852	membrane alanyl aminopeptidase (EC 3.4.11.2) - rat, 965 aa.	6..962 2..920	339/975 (34%) 514/975 (51%)	e-155
P15684	Aminopeptidase N (EC 3.4.11.2) (Microsomal aminopeptidase) - Rattus norvegicus (Rat), 964 aa.	6..962 1..919	339/975 (34%) 514/975 (51%)	e-155
A53984	membrane alanyl aminopeptidase (EC 3.4.11.2) - pig, 963 aa.	6..968 2..924	341/983 (34%) 523/983 (52%)	e-154

PFam analysis predicts that the NOV10a protein contains the domains shown in the Table 10F.

Table 10F. Domain Analysis of NOV10a			
Pfam Domain	NOV10a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Peptidase_M1	98..509	150/451 (33%) 323/451 (72%)	9.1e-144

5

140

EXAMPLE 11.

The NOV11 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 11A.

Table 11A. NOV11 Sequence Analysis		
	SEQ ID NO: 37	818 bp
NOV11a, CG95430-01 DNA Sequence	TCCCTCTTTCAGTTCAGAGTCTGTCATCTGAACCATGAGGATCTGGTGGCTTCTGCTT GCCATTGAAATCTGCACAGGGAACATAAACTCACAGGACACCTGCAGGCAAGGGCACC CTGGAATCCCTGGGAACCCCGGTCACAATGGTCTGCCTGGAAGAGATGGACGAGACGG AGCGAAGGGTGACAAAGGCGATGCAGGTAAGCCTGGTCCCAAAGGAGAAGCTGGACCC ACGGGGCCCCAGGGTGAGCCAGGAGTCCGGGGAATAAGAGGCTGGAAAGGAGATCGAG GAGAGAAAGGGAAAATCGGTGAGACTCTAGTCTTGCCAAAAAGTGCTTTCACTGTGGG GCTCACGGTGCTGAGCAAGTTTCCTTCTTCAGATATGCCCATTAAATTTGATAAGATC CTGTATAACGAATTCAACCATTATGATACAGCAGCGGGGAAATTCACGTGCCACATTG CTGGGGTCTATTACTTCACCTACCACATCACTGTTTTCTCCAGAAATGTTCAGGTGTC TTTGGTCAAAAATGGAGTAAAAATACTGCACACCAAAGATGCTTACATGAGCTCTGAG GACCAGGCCTCTGGCGGCATTGTCCTGCAGCTGAAGCTCGGGGATGAGGTGTGGCTGC AGGTGACAGGAGGAGAGAGGTTCAATGGCTTGTTTGCTGATGAGGACGATGACACAAC TTTCACAGGGTTCCTTCTGTTCAGCAGCCCGTGACAGAGGAGAGTTTAAAAATCCGCC ACACCATCCATCAGAATCAGCTTGGGATGAACTTATTCAGATGGTTTTACTTTATTAA TTCCTC
	ORF Start: ATG at 35	ORF Stop: TGA at 728
	SEQ ID NO: 38	231 aa	MW at 24946.0kD
NOV11a, CG95430-01 Protein Sequence	MRIWWLLLAIEICTGNINSQDTCRQGHPGIPGNPGHNGLPGRDGRDGAKGDKGDAGKP GPKGEAGPTGPQGEPGVRGIRGWKGDRGEKGKIGETLVLPKSAFTVGLTVLSKFPSSD MPIKFDKILYNEFNHYDTAAGKFTCHIAGVYYFTYHITVFSRNVQVSLVKNGVKILHT KDAYMSSEDQASGGIVLQLKLGDEVWLQVTGGERFNGLFADEDDDTTFTGFLLFSSP	
	SEQ ID NO: 39	954 bp
NOV11b, CG95430-02 DNA Sequence	GGATCCCAGGACACCTGCAGGCAAGGGCACCCTGGGATCCCTGGGAACCCCGGTCACA ATGGTCTGCCTGGAAGAGATGGACGAGACGGAGCGAAGGGTGACAAAGGCGATGCAGG AGAACCAGGACGTCCTGGCAGCCCGGGGAAGGATGGGACGAGTGGAGAGAAGGGAGAA CGAGGAGCAGATGGAAAAGTTGAAGCAAAAGGCATCAAAGGTGATCAAGGCTCAAGAG GATCCCCAGGAAAACATGGCCCCAAGGGGCTTGCAGGGCCCATGGGAGAGAAAGGCCT CCGAGGAGAGACTGGGCCTCAGGGGCAGAAGGGGAATAAGGGTGACGTGGGTCCCACT GGTCCTGAGGGGCCAAGGGGCAACATTGGGCCTTTGGGCCCAACTGGTTTACCGGGCC CCATGGGCCCTATTGGAAAGCCTGGTCCCAAGGGAGAAGCTGGACCCACGGGGCCCCA GGGTGAGCCAGGAGTCCGGGGAATAAGAGGCTGGAAAGGAGATCGAGGAGAGAAAGGG AAAATCGGTGAGACTCTAGTCTTGCCAAAAAGTGCTTTCACTGTGGGGCTCACGGTGC TGAGCAAGTTTCCTTCTTCAGATGTGCCCATTAAATTTGATAAGATCCTGTATAACGA ATTCAACCATTATGATACAGCAGCGGGGAAATTCACGTGCCACATTGCTGGGGTCTAT TACTTCACCTACCACATCACTGTTTTCTCCAGGAATGTTCAGGTGTCTTTGGTCAAAA ATGGAGTAAAAATACTGCACACCAAAGATGCTTACATGAGCTCTGAGGACCAGGCCTC TGGCGGCATTGTCCTGCAGCTGAAGCTCGGGGATGAGGTGTGGCTGCAGGTGACAGGA GGAGAGAGGTTCAATGGCTTGTTTGCTGATGAGGACGATGACACAACTTTCACAGGGT TCCTTCTGTTCAGCAGCCCGCTCGAG
	ORF Start: at 7	ORF Stop: at 949
	SEQ ID NO: 40	314 aa	MW at 32420.0kD
NOV11b, CG95430-02 Protein Sequence	QDTCRQGHPGIPGNPGHNGLPGRDGRDGAKGDKGDAGEPGRPGSPGKDGTSGEKGERG ADGKVEAKGIKGDQGSRGSPGKHGPKGLAGPMGEKGLRGETGPQGQKGNKGDVGPTGP EGPRGNIGPLGPTGLPGPMGPIGKPGPKGEAGPTGPQGEPGVRGIRGWKGDRGEKGKI GETLVLPKSAFTVGLTVLSKFPSSDVPIKFDKILYNEFNHYDTAAGKFTCHIAGVYYF	

141

	TYHITVFSRNVQVSLVKNGVKILHTKDAYMSSEDQASGGIVLQLKLGDEVWLQVTGGE RFNGLFADEDDDTTFTGFLLFSSP	
	SEQ ID NO: 41	405 bp
NOV11c, CG95430-03 DNA Sequence	GGATCCGCTTTCACTGTGGGGCTCACGGTGCTGAGCAAGTTTCCTTCTTCAGATATGC CCATTAAATTTGATAAGATCCTGTATAACGAATTCAACCATTATGATACAGCAGCGGG GAAATTCACGTGCCACATTGCTGGGGTCTATTACTTCACCTACCACATCACTGTTTTC TCCAGGAATGTTCAGGTGTCTTTGGTCAAAAATGGAGTAAAAATACTGCACACCAAAG ATGCTTACATGAGCTCTGAGGACCAGGCCTCTGGCGGCATTGTCCTGCAGCTGAAGCT CGGGGATGAGGTGTGGCTGCAGGTGACAGGAGGAGAGAGGTTCAATGGCTTGTTTGCT GATGAGGACGATGACACAACTTTCACAGGGTTCCTTCTGTTCAGCAGCCCGCTCGAG	
	ORF Start: at 7	ORF Stop: at 400
	SEQ ID NO: 42	131 aa	MW at 14607.4kD
NOV11c, CG95430-03 Protein Sequence	AFTVGLTVLSKFPSSDMPIKFDKILYNEFNHYDTAAGKFTCHIAGVYYFTYHITVFSR NVQVSLVKNGVKILHTKDAYMSSEDQASGGIVLQLKLGDEVWLQVTGGERFNGLFADE DDDTTFTGFLLFSSP	
	SEQ ID NO: 43	1026 bp
NOV11d, CG95430-04 DNA Sequence	TCTGTCATCTGAACCATGAGGATCTGGTGGTTTCTGCTTGCCATTGAAATCTGCACAG GGAACATAAACTCACAGGACACCTGCAGGCAAGGGCACCCTGGCATCCCTGGGAACCC CGGTCACAATGGTCTGTCTGGAAGAGATGGACGAGACGGAGCGAAGGGTGACAAAGGC GATGCAGGAGAACCAGGACGTCCTGGCAGCCCGGGGAAGGATGGGACGAGTGGAGAGA AGGGAGAACGAGGAGCAGATGGAAAAGTTGAAGCAAAAGGCATCAAAGGTGATCAAGG CTCAAGAGGATCCCCAGGAAAACATGGCCCCAAGGGGCTTGCAGGGCCCATGGGAGAG AAGGGCCTCCGAGGAGAGACTGGGCCTCAGGGGCAGAAGGGGAATAAGGGTGACGTGG GTCCCACTGGTCCTGAGGGGCCAAGGGGCAACATTGGGCCTTTGGGCCCAACTGGTTT ACCGGGCCCCATGGGCCCTATTGGAAAGCCTGGTCCCAAAGGAGAAGCTGGACCCACG GGGCCCCAGGGTGAGCCAGGAGTCCGGGGAATAAGAGGCTGGAAAGGAGATCGAGGAG AGAAAGGGAAAATCGGTGAGACTCTAGTCTTGCCAAAAAGTGCTTTCACTGTGGGGCT CACGGTGCTGAGCAAGTTTCCTTCTTCAGATATGCCCATTAAATTTGATAAGATCCTG TATAACGAATTCAACCATTATGATACAGCAGCGGGGAAATTCACGTGCCACATTGCTG GGGTCTATTACTTCACCTACCACATCACTGTTTTCTCCAGGAATGTTCAGGTGTCTTT GGTCAAAAATGGAGTAAAAATACTGCACACCAAAGATGCTTACATGAGCTCTGAGGAC CAGGCCTCTGGCGGCATTGTCCTGCAGCTGAAGCTCGGGGATGAGGTGTGGCTGCAGG TGACAGGAGGAGAGAGGTTCAATGGCTTGTTTGCTGATGAGGACGATGACACAACTTT CACAGGGTTCCTTCTGTTCAGCAGCCAGTGACAGAGGAGA	
	ORF Start: ATG at 16	ORF Stop: TGA at 1015
	SEQ ID NO: 44	333 aa	MW at 34735.7kD
NOV11d, CG95430-04 Protein Sequence	MRIWWFLLAIEICTGNINSQDTCRQGHPGIPGNPGHNGLSGRDGRDGAKGDKGDAGEP GRPGSPGKDGTSGEKGERGADGKVEAKGIKGDQGSRGSPGKHGPKGLAGPMGEKGLRG ETGPQGQKGNKGDVGPTGPEGPRGNIGPLGPTGLPGPMGPIGKPGPKGEAGPTGPQGE PGVRGIRGWKGDRGEKGKIGETLVLPKSAFTVGLTVLSKFPSSDMPIKFDKILYNEFN HYDTAAGKFTCHIAGVYYFTYHITVFSRNVQVSLVKNGVKILHTKDAYMSSEDQASGG IVLQLKLGDEVWLQVTGGERFNGLFADEDDDTTFTGFLLFSSQ	
	SEQ ID NO: 45	889 bp
NOV11e, CG95430-06 DNA Sequence	TCTGTCATCTGAACCATGAGGATCTGGTGGTTTCTGCTTGCCATTGAAATCTGCACAG GGAACATAAACTCACAGGACACCTGCAGGCAAGGGCACCCTGGCATCCCTGGGAACCC CGGTCACAATGGTCTGTCTGGAAGAGATGGACGAGACGGAGCGAAGGGTGACAAAGGC GATGCAGGAGAACCAGGACGTCCTGGCAGCCCGGGGAAGGATGGGACGAGTGGAGAGA AGGGAGAACGAGGAGCAGATGGAAAAGTTGAAGCAAAAGGCATCAAAGGTGATCAAGG CTCAAGAGGATCCCCAGGAAAACATGGCCCCAAGGGGCTTGCAGGGCCCATGGGAGAG AAGGGCCTCCGAGGAGAGACTGGGCCTCAGGGGCAGAAGGGGAATAAGGGTGAGCCAG GAGTCCGGGGAATAAGAGGCTGGAAAGGAGATCGAGGAGAGAAAGGGAAAATCGGTGA GACTCTAGTCTTGCCAAAAAGTGCTTTCACTGTGGGGCTCACGGTGCTGAGCAAGTTT	

	CCTTCTTCAGATGTGCCCATTAAATTTGATAAGATCCTGTATAACGAATTCAACCATT ATGATACAGCAGCGGGGAAATTCACGTGCCACATTGCTGGGGTCTATTACTTCACCTA CCACATCGCTGTTTTCTCCAGCAATGTTCAGGTGTCTTTGGTCAAAAATGGAGTAAAA ATACTGCACACCAAAGATGCTTACATGAGCTCTGAGGACCAGGCCTCTGGCGGCATTG TCCTGCAGCTGAAGCTCGGGGATGAGGTGTGGCTGCAGGTGACAGGAGGAGAGAGGTT CAATGGCTTGTTTGCTGATGAGGACGATGACACAACTTTCACAGGGTTCCTTCTGTTC AGCAGCCAG**TGA**CAGAGGA

	ORF Start: ATG at 16		ORF Stop: TGA at 880
	SEQ ID NO: 46		288 aa	MW at 30497.9kD
NOV11e, CG95430-06 Protein Sequence	MRIWWFLLAIEICTGNINSQDTCRQGHPGIPGNPGHNGLSGRDGRDGAKGDKGDAGEP GRPGSPGKDGTSGEKGERGADGKVEAKGIKGDQGSRGSPGKHGPKGLAGPMGEKGLRG ETGPQGQKGNKGEPGVRGIRGWKGDRGEKGKIGETLVLPKSAFTVGLTVLSKFPSSDV PIKFDKILYNEFNHYDTAAGKFTCHIAGVYYFTYHIAVFSSNVQVSLVKNGVKILHTK DAYMSSEDQASGGIVLQLKLGDEVWLQVTGGERFNGLFADEDDDTTFTGFLLFSSQ		

	SEQ ID NO: 47	405 bp
NOV11f, 175184045 DNA Sequence	GGATCCGCTTTCACTGTGGGGCTCACGGTGCTGAGCAAGTTTCCTTCTTCAGATATGC CCATTAAATTTGATAAGATCCTGTATAACGAATTCAACCATTATGATACAGCAGCGGG GAAATTCACGTGCCACATTGCTGGGGTCTATTACTTCACCTACCACATCACTGTTTTC TCCAGGAATGTTCAGGTGTCTTTGGTCAAAAATGGAGTAAAAATACTGCACACCAAAG ATGCTTACATGAGCTTTGAGGACCAGGCCTCTGGCGGCATTGTCCTGCAGCTGAAGCT CGGGGATGAGGTGTGGCTGCAGGTGACAGGAGGAGAGAGGTTCAATGGCTTGTTTGCT GATGAGGACGATGACACAACTTTCACAGGGTTCCTTCTGTTCAGCAGCCCGCTCGAG

	ORF Start: at 1	ORF Stop: end of sequence
	SEQ ID NO: 48	135 aa	MW at 15053.9kD

	SEQ ID NO: 48	135 aa	MW at 15053.9kD
NOV11f, 175184045 Protein Sequence	GSAFTVGLTVLSKFPSSDMPIKFDKILYNEFNHYDTAAGKFTCHIAGVYYFTYHITVF SRNVQVSLVKNGVKILHTKDAYMSFEDQASGGIVLQLKLGDEVWLQVTGGERFNGLFA DEDDDTTFTGFLLFSSPLE		

	SEQ ID NO: 49	405 bp
NOV11g, 175184049 DNA Sequence	GGATCCGCTTTCACTGTGGGGCTCACGGTGCTGAGCAAGTTTCCTTCTTCAGATATGC CCATTAAATTTGATAAGATCCTGTATAACGAATTCAACCATTATGATACAGCAGCGGG GAAATTCACGTGCCACATTGCTGGGGTCTATTACTTCACCTACCACATCACTGTTTTC TCCAGAAATGTTCAGGTGTCTTTGGTCAAAAATGGAGTAAAAATACTGCACACCAAAG ATGCTTACATGAGCTCTGAGGACCAGGCCTCTGGCGGCATTGTCCTGCAGCTGAAGCT CGGGGATGAGATGTGGCTGCAGGTGACAGGAGGAGAGAGGTTCAATGGCTTGTTTGCT GATGAGGACGATGACACAACTTTCACAGGGTTCCTTCTGTTCAGCAGCCCGCTCGAG

	ORF Start: at 1	ORF Stop: end of sequence
	SEQ ID NO: 50	135 aa	MW at 15025.8kD

	SEQ ID NO: 50	135 aa	MW at 15025.8kD
NOV11g, 175184049 Protein Sequence	GSAFTVGLTVLSKFPSSDMPIKFDKILYNEFNHYDTAAGKFTCHIAGVYYFTYHITVF SRNVQVSLVKNGVKILHTKDAYMSSEDQASGGIVLQLKLGDEMWLQVTGGERFNGLFA DEDDDTTFTGFLLFSSPLE		

	SEQ ID NO: 51	405 bp
NOV11h, 175184053 DNA Sequence	GGATCCGCTTTCACTGTGGGGCTCACGGTGCTGAGCAAGTTTCCTTCTTCAGATATGC CCATTAAATTTGATAAGATCCTGTATAACGAATTCAACCATTATGATACAGCAGCGGG GAAATTCACGTGCCACATTGCTGGGGTCTATTACTTCACCTACCACATCACTGTTTTC TCCAGGAATGTTCAGGTGTCTTTGGTCAAAAATGGAGTAAAAATACTGCACACCAAAG ATGCTTACATGAGCTCTGAGGACCAGGCCTCTGGCGGCATTGTCCTGCAGCTGAAGCT CGGGGATGAGGTGTGGCTGCAGGTGACAGGAGGAGAGAGGTTCAATGGCTTGTTTGCT GATGAGGACGATGACACAACTTTCACAGGGTTCCTTCTGTTCAGCAGCCCGCTCGAG

	ORF Start: at 1	ORF Stop: end of sequence

	SEQ ID NO: 52	135 aa	MW at 14993.8kD
NOV11h, 175184053 Protein Sequence	GSAFTVGLTVLSKFPSSDMPIKFDKILYNEFNHYDTAAGKFTCHIAGVYYFTYHITVF SRNVQVSLVKNGVKILHTKDAYMSSEDQASGGIVLQLKLGDEVWLQVTGGERFNGLFA DEDDDTTFTGFLLFSSPLE		
	SEQ ID NO: 53	954 bp	
NOV11i, 175070796 DNA Sequence	**GGA**TCCCAGGACACCTGCAGGCAAGGGCACCCTGGGATCCCTGGGAACCCCGGTCACA ATGGTCTGCCTGGAAGAGATGGACGAGACGGAGCGAAGGGTGACAAAGGCGATGCAGG AGAACCAGGACGTCCTGGCAGCCCGGGGAAGGATGGGACGAGTGGAGAGAAGGGAGAA CGAGGAGCAGATGGAAAAGTTGAAGCAAAAGGCATCAAAGGTGATCAAGGCTCAAGAG GATCCCCAGGAAAACATGGCCCCAAGGGGCTTGCAGGGCCCATGGGAGAGAAAGGCCT CCGAGGAGAGACTGGGCCTCAGGGGCAGAAGGGGAATAAGGGTGACGTGGGTCCCACT GGTCCTGAGGGGCCAAGGGGCAACATTGGGCCTTTGGGCCCAACTGGTTTACCGGGCC CCATGGGCCCTATTGGAAAGCCTGGTCCCAAGGGAGAAGCTGGACCCACGGGGCCCCA GGGTGAGCCAGGAGTCCGGGGAATAAGAGGCTGGAAAGGAGATCGAGGAGAGAAAGGG AAAATCGGTGAGACTCTAGTCTTGCCAAAAAGTGCTTTCACTGTGGGGCTCACGGTGC TGAGCAAGTTTCCTTCTTCAGATGTGCCCATTAAATTTGATAAGATCCTGTATAACGA ATTCAACCATTATGATACAGCAGCGGGGAAATTCACGTGCCACATTGCTGGGGTCTAT TACTTCACCTACCACATCACTGTTTTCTCCAGGAATGTTCAGGTGTCTTTGGTCAAAA ATGGAGTAAAAATACTGCACACCAAAGATGCTTACATGAGCTCTGAGGACCTGGCCTC TGGCGGCATTGTCCTGCAGCTGAAGCTCGGGGATGAGGTGTGGCTGCAGGTGACAGGA GGAGAGAGGTTCAATGGCTTGTTTGCTGATGAGGACGATGACACAACTTTCACAGGGT TCCTTCTGTTCAGCAGCCCGCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 54	318 aa	MW at 32791.4kD
NOV11i, 175070796 Protein Sequence	GSQDTCRQGHPGIPGNPGHNGLPGRDGRDGAKGDKGDAGEPGRPGSPGKDGTSGEKGE RGADGKVEAKGIKGDQGSRGSPGKHGPKGLAGPMGEKGLRGETGPQGQKGNKGDVGPT GPEGPRGNIGPLGPTGLPGPMGPIGKPGPKGEAGPTGPQGEPGVRGIRGWKGDRGEKG KIGETLVLPKSAFTVGLTVLSKFPSSDVPIKFDKILYNEFNHYDTAAGKFTCHIAGVY YFTYHITVFSRNVQVSLVKNGVKILHTKDAYMSSEDLASGGIVLQLKLGDEVWLQVTG GERFNGLFADEDDDTTFTGFLLFSSPLE		
	SEQ ID NO: 55	954 bp	
NOV11j, 175070804 DNA Sequence	**GGA**TCCCAGGACACCTGCAGGCAAGGGCACCCTGGGATCCCTGGGAACCCCGGTCACA ATGGTCTGCCTGGAAGAGATGGACGAGACGGAGCGAAGGGTGACAAAGGCGATGCAGG AGAACCAGGACGTCCTGGCAGCCCGGGGAAGGATGGGACGAGTGGAGAGAAGGGAGAA CGAGGAGCAGATGGAAAAGTTGAAGCAAAAGGCATCAAAGGTGATCAAGGCTCAAGAG GATCCCCAGGAAAACATGGCCCCAAGGGGTTTGCAGGGCCCATGGGAGAGAAAGGCCT CCGAGGAGAGACTGGGCCTCAGGGGCAGAAGGGGAATAAGGGTGACGTGGGTCCCACT GGTCCTGAGGGGCCAAGGGGCAACATTGGGCCTTTGGGCCCAACTGGTTTACCGGGCC CCATGGGCCCTATTGGAAAGCCTGGTCCCAAGGGAGAAGCTGGACCCACGGGGCCCCA GGGTGAGCCAGGAGTCCGGGGAATAAGAGGCTGGAAAGGAGATCGAGGAGAGAAAGGG AAAATCGGTGAGACTCTAGTCTTGCCAAAAAGTGCTTTCACTGTGGGGCTCACGGTGC TGAGCAAGTTTCCTTCTTCAGATGTGCCCATTAAATTTGATAAGATCCTGTATAACGA ATTCAACCATTATGATACAGCAGCGGGGAAATTCACGTGCCACATTGCTGGGGTCTAT TACTTCACCTACCACATCACTGTTTTCTCCAGGAATGTTCAGGTGTCTTTGGTCAAAA ATGGAGTAAAAATACTGCACACCAAAGATGCTTACATGAGCTCTGAGGACCAGGCCTC TGGCGGCATTGTCCTGCAGCTGAAGCTCGGGGATGAGGTGTGGCTCCAGGTGACAGGA GGAGAGAGGTTCAATGGCTTGTTTGCTGATGAGGACGATGACACAACTTTCACAGGGT TCCTTCTGTTCAGCAGCCCGCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 56	318 aa	MW at 32840.4kD
NOV11j, 175070804	GSQDTCRQGHPGIPGNPGHNGLPGRDGRDGAKGDKGDAGEPGRPGSPGKDGTSGEKGE RGADGKVEAKGIKGDQGSRGSPGKHGPKGFAGPMGEKGLRGETGPQGQKGNKGDVGPT GPEGPRGNIGPLGPTGLPGPMGPIGKPGPKGEAGPTGPQGEPGVRGIRGWKGDRGEKG		

Protein Sequence	KIGETLVLPKSAFTVGLTVLSKFPSSDVPIKFDKILYNEFNHYDTAAGKFTCHIAGVY YFTYHITVFSRNVQVSLVKNGVKILHTKDAYMSSEDQASGGIVLQLKLGDEVWLQVTG GERFNGLFADEDDDTTFTGFLLFSSPLE	
	SEQ ID NO: 57	954 bp
NOV11k, 175070808 DNA Sequence	GGATCCCAGGACACCTGCAGGCAAGGGCACCCTGGGATCCCTGGGAACCCCGGTCACA ATGGTCTGCCTGGAAGAGATGGACGAGACGGAGCGAAGGGTGACAAAGGCGATGCAGG AGAACCAGGACGTCCTGGCAGCCCGGGGAAGGATGGGACGAGTGGAGAGAAGGGAGAA CGAGGAGCAGATGGAAAAGTTGAAGCAAAAGGCATCAAAGGTGATCAAGGCTCAAGAG GATCCCCAGGAAAACATGGCCCCAAGGGGCTTGCAGGGCCCATGGGAGAGAAAGGCCT CCGAGGAGAGACTGGGCCTCAGGGGCAGAAGGGGAATAAGGGTGACGTGGGTCCCACT GGTCCTGAGGGGCCAAGGGGCAACATTGGGCCTTTGGGCCCAACTGGTTTACCGGGCC CCATGGGCCCTATTGGAAAGCCTGGTCCCAAGGGAGAAGCTGGACCCACGGGGCCCCA GGGTGAGCCAGGAGTCCGGGGAATAAGAGGCTGGAAAGGAGATCGAGGAGAGAAAGGG AAAATCGGTGAGACTCTAGTCTTGCCAAAAAGTGCTTTCACTGTGGGGCTCACGGTGC TGAGCAAGTTTCCTTCTTCAGATGTGCCCATTAAATTTGATAAGATCCTGTATAACGA ATTCAACCATTATGATACAGCAGCGGGGAAATTCACGTGCCACATTGCTGGGGTCTAT TACTTCACCTACCACATCACTGTTTTCTCCAGGAATGTTCAGGTGTCTTTGGTCAAAA ATGGAGTAAAAATACTGCACACCAAAGATGCTTACATGAGCTCTGAGGACCAGGCCTC TGGCGGCATTGTCCTGCAGCTGAAGCTCGGGGATGAGGTGTGGCTGCAGGTGACAGGA GGAGAGAGGTTCAATGGCTTGTTTGCTGATGAGGACGATGACACAACTTTCACAGGGT TCCTTCTGTTCAGCAGCCCGCTCGAG	
	ORF Start: at 1	ORF Stop: end of sequence
	SEQ ID NO: 58	318 aa	MW at 32806.4kD
NOV11k, 175070808 Protein Sequence	GSQDTCRQGHPGIPGNPGHNGLPGRDGRDGAKGDKGDAGEPGRPGSPGKDGTSGEKGE RGADGKVEAKGIKGDQGSRGSPGKHGPKGLAGPMGEKGLRGETGPQGQKGNKGDVGPT GPEGPRGNIGPLGPTGLPGPMGPIGKPGPKGEAGPTGPQGEPGVRGIRGWKGDRGEKG KIGETLVLPKSAFTVGLTVLSKFPSSDVPIKFDKILYNEFNHYDTAAGKFTCHIAGVY YFTYHITVFSRNVQVSLVKNGVKILHTKDAYMSSEDQASGGIVLQLKLGDEVWLQVTG GERFNGLFADEDDDTTFTGFLLFSSPLE	
	SEQ ID NO: 59	954 bp
NOV11l, 175070812 DNA Sequence	GGATCCCAGGACACCTGCAGGCAAGGGCACCCTGGGATCCCTGGGAACCCCGGTCACA ATGGTCTGCCTGGAAGAGATGGACGAGACGGAGCGAAGGGTGACAAAGGCGATGCAGG AGAACCAGGACGTCCTGGCAGCCCGGGGAAGGATGGGACGAGTGGAGAGAAGGGAGAA CGAGGAGCAGATGGAAAAGTTGAAGCAAAAGGCATCAAAGGTGATCAAGGCTCAAGAG GATCCCCAGGAAAACATGGCCCCAAGGGGCTTGCAGGGCCCATGGGAGAGAAAGGCCT CCGAGGAGAGACTGGGCCTCAGGGGCAGAAGGGGAATAAGGGTGACGTGGGTCCCACT GGTCCTGAGGGGCCAAGGGGCAACATTGGGCCTTTGGGCCCAACTGGTTTACCGGGCC CCATGGGCCCTATTGGAAAGCCTGGTCCCAAGGGAGAAGCTGGACCCACGGGGCCCCA GGGTGAGCCAGGAGTCCGGGGAATAAGAGGCTGGAAAGGAGATCGAGGAGAGAAAGGG AAAATCGGTGAGACTCTAGTCTTGCCAAAAAGTGCTTTCACTGTGGGGCTCACGGTGC TGAGCAAGTTTCCTTCTTCAGATGTGCCCATTAAATTTGATAAGATCCTGTATAACGA ATTCAACCATTATGATACAGCAGCGGGGAAATTCACGTGCCACATTGCTGGGGTCTAT TACTTCACCTACCACATCACTGTTTTCTCCAGGAATGTTCAGGTGTCTTTGGTCAAAA ATGGAGTAAAAATACTGCACACCAAAGATGCTTACATGAGCTCTGAGGACCAGGTCTC TGGCGGCATTGTCCTGCAGCTGAAGCTCGGGGATGAGGTGTGGCTGCAGGTGACAGGA GGAGAGAGGTTCAATGGCTTGTTTGCTGATGAGGACGATGACACAACTTTCACAGGGT TCCTTCTGTTCAGCAGCCCGCTCGAG	
	ORF Start: at 1	ORF Stop: end of sequence
	SEQ ID NO: 60	318 aa	MW at 32834.5kD
NOV11l, 175070812 Protein Sequence	GSQDTCRQGHPGIPGNPGHNGLPGRDGRDGAKGDKGDAGEPGRPGSPGKDGTSGEKGE RGADGKVEAKGIKGDQGSRGSPGKHGPKGLAGPMGEKGLRGETGPQGQKGNKGDVGPT GPEGPRGNIGPLGPTGLPGPMGPIGKPGPKGEAGPTGPQGEPGVRGIRGWKGDRGEKG KIGETLVLPKSAFTVGLTVLSKFPSSDVPIKFDKILYNEFNHYDTAAGKFTCHIAGVY YFTYHITVFSRNVQVSLVKNGVKILHTKDAYMSSEDQVSGGIVLQLKLGDEVWLQVTG GERFNGLFADEDDDTTFTGFLLFSSPLE	

	SEQ ID NO: 61	954 bp	
NOV11m, 175070828 DNA Sequence	GGATCCCAGGACACCTGCAGGCAAGGGCACCCTGGGATCCCTGGGAACCCCGGTCACA ATGGTCTGCCTGGAAGAGATGGACGAGACGGAGCGAAGGGTGACAAAGGCGATGCAGG AGAACCAGGACGTCCTGGCAGCCCGGGGAAGGATGGGACGAGTGGAGAGAAGGGAGAA CGAGGAGCAGATGGAAAAGTTGAAGCAAAAGGCATCAAAGGTGATCAAGGCTCAAGAG GATCCCCAGGAAAACATGGCCCCAAGGGGCTTGCAGGGCCCATGGGAGAGAAGGGCCT CCGAGGAGAGACTGGGCCTCAGGGGCAGAAGGGGAATAAGGGTGACGCGGGTCCCACT GGTCCTGAGGGGCCAAGGGGCAACATTGGGCCTTTGGGCCCAACTGGTTTACCGGGCC CCATGGGCCCTATTGGAAAGCCTGGTCCCAAAGGAGAAGCTGGACCCACGGGGCCCCA GGGTGAGCCAGGAGTCCAGGGAATAAGAGGCTGGAAAGGAGATCGAGGAGAGAAAGGG AAAATCGGTGAGACTCTAGTCTTGCCAAAAAGTGCTTTCACTGTGGGGCTCACGGTGC TGAGCAAGTTTCCTTCTTCAGATATGCCCATTAAATTTGATAAGATCCTGTATAACGA ATTCAACCATTATGATACAGCAGCGGGGAAATTCACGTGCCACATTGCTGGGGTCTAT TACTTCACCTACCACATCACTGTTTTCTCCAGGAATGTTCAGGTGTCTTTGGTCAAAA ATGGAGTAAAAATACTGCACACCAAAGATGCTTACATGAGCTCTGAGGACCAGGCCTC TGGCGGCATTGTCCTGCAGCTGAAGCTCGGGGATGAGGTGTGGCTGCAGGTGACAGGA GGAGAGAGGTTCAATGGCTTGTTTGCTGATGAGGACGATGACACAACTTTCACAGGGT TCCTTCTGTTCAGCAGCCCGCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 62	318 aa	MW at 32782.4kD
NOV11m, 175070828 Protein Sequence	GSQDTCRQGHPGIPGNPGHNGLPGRDGRDGAKGDKGDAGEPGRPGSPGKDGTSGEKGE RGADGKVEAKGIKGDQGSRGSPGKHGPKGLAGPMGEKGLRGETGPQGQKGNKGDAGPT GPEGPRGNIGPLGPTGLPGPMGPIGKPGPKGEAGPTGPQGEPGVQGIRGWKGDRGEKG KIGETLVLPKSAFTVGLTVLSKFPSSDMPIKFDKILYNEFNHYDTAAGKFTCHIAGVY YFTYHITVFSRNVQVSLVKNGVKILHTKDAYMSSEDQASGGIVLQLKLGDEVWLQVTG GERFNGLFADEDDDTTFTGFLLFSSPLE		
	SEQ ID NO: 63	954 bp	
NOV11n, 175070836 DNA Sequence	GGATCCCAGGACACCTGCAGGCAAGGGCACCCTGGGATCCCTGGGAACCCCGGTCACA ATGGTCTGCCTGGAAGAGATGGACGAGACGGAGCGAAGGGTGACAAAGGCGATGCAGG AGAACCAGGACGTCCTGGCAGCCCGGGGAAGGATGGGACGAGTGGAGAGAAGGGAGAA CGAGGAGCAGATGGAAAAGTTGAAGCAAAAGGCATCAAAGGTGATCAAGGCTCAAGAG GATCCCCAGGAAAACATGGCCCCAAGGGGCTTGCAGGGCCCATGGGAGAGAAGGGCCT CCGAGGAGAGACTGGGCCTCAGGGGCAGAAGGGGAATAAGGGTGACGTGGGTCCCACT GGTCCTGAGGGGCCAAGGGGCAACATTGGGCCTTTGGGCCCAACTGGTTTACCGGGCC CCATGGGCCCTATTGGAAAGCCTGGTCCCAAAGGAGAAGCTGGACCCACGGGGCCCCA GGGTGAGCCAGGAGTCCAGGGAATAAGAGGCTGGAAAGGAGATCGAGGAGAGAAAGGG AAAATCGGTGAGACTCTAGTCTTGCCAAAAAGTGCTTTCACTGTGGGGCTCACGGTGC TGAGCAAGTTTCCTTCTTCAGATATGCCCATTAAATTTGATAAGATCCTGTATAACGA ATTCAACCATTATGATACAGCAGCGGGGAAATTCACGTGCCACATTGCTGGGGTCTAT TACTTCACCTACCACATCACTGTTTTCTCCAGGAATGTTCAGGTGTCTTTGGTCAAAA ATGGAGTAAAAATACTGCACACCAAAGATGCTTACATGAGCTCTGAGGACCAGGCCTC TGGCGGCATTGTCCTGCAGCTGAAGCTCGGGGATGAGGTGTGGCTGCAGGTGACAGGA GGAGAGAGGTTCAATGGCTTGTTTGCTGATGAGGACGATGACACAACTTTCACAGGGT TCCTTCTGTTCAGCAGCCCGCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 64	318 aa	MW at 32810.4kD
NOV11n, 175070836 Protein Sequence	GSQDTCRQGHPGIPGNPGHNGLPGRDGRDGAKGDKGDAGEPGRPGSPGKDGTSGEKGE RGADGKVEAKGIKGDQGSRGSPGKHGPKGLAGPMGEKGLRGETGPQGQKGNKGDVGPT GPEGPRGNIGPLGPTGLPGPMGPIGKPGPKGEAGPTGPQGEPGVQGIRGWKGDRGEKG KIGETLVLPKSAFTVGLTVLSKFPSSDMPIKFDKILYNEFNHYDTAAGKFTCHIAGVY YFTYHITVFSRNVQVSLVKNGVKILHTKDAYMSSEDQASGGIVLQLKLGDEVWLQVTG GERFNGLFADEDDDTTFTGFLLFSSPLE		
	SEQ ID NO: 65	954 bp	
NOV11o,	GGATCCCAGGACACCTGCAGGCAAGGGCACCCTGGGATCCCTGGGAACCCCGGTCACA		

175070840 DNA Sequence	ATGGTCTGCCTGGAAGAGATGGACGAGACGGAGCGAAGGGTGACAAAGGCGATGCAGG AGAACCAGGACGTCCTGGCAGCCCGGGGAAGGATGGGACGAGTGGAGAGAAGGGAGAA CGAGGAGCAGATGGAAAAGTTGAAGCAAAAGGCATCAAAGGTGATCAAGGCTCAAGAG GATCCCCAGGAAAACATGGCCCCAAGGGGCTTGCAGGGCCCATGGGAGAGAAGGGCCT CCGAGGGGAGACTGGGCCTCAGGGGCAGAAGGGGAATAAGGGTGACGTGGGTCCCACT GGTCCTGAGGGGCCAAGGGGCAACATTGGGCCTTTGGGCCCAACTGGTTTACCGGGCC CCATGGGCCCTATTGGAAAGCCTGGTCCCAAAGGAGAAGCTGGACCCACGGGGCCCCA GGGTGAGCCAGGAGTCCGGGGAATAAGAGGCTGGAAAGGAGATCGAGGAGAGAAAGGG AAAATCGGTGAGACTCTAGTCTTGCCAAAAAGTGCTTTCACTGTGGGGCTCACGGTGC TGAGCAAGTTTCCTTCTTCAGATATGCCCATTAAATTTGATAAGATCCTGTATAACGA ATTCAACCATTATGATACAGCAGCGGGGAAATTCACGTGCCACATTGCTGGGGTCTAT TACTTCACCTACCACATCACTGTTTTCTCCAGAAATGTTCAGGTGTCTTTGGTCAAAA ATGGAGTAAAAATACTGCACACCAAAGATGCTTACATGAGCTCTGAGGACCAGGCCTC TGGCGGCATTGTCCTGCAGCTGAAGCTCGGGGATGAGATGTGGCTGCAGGTGACAGGA GGAGAGAGGTTCAATGGCTTGTTTGCTGATGAGGACGATGACACAACTTTCACAGGGT TCCTTCTGTTCAGCAGCCCGCTCGAG
	ORF Start: at 1 **ORF Stop: end of sequence**
	SEQ ID NO: 66 **318 aa** **MW at 32870.5kD**
NOV11o, 175070840 Protein Sequence	GSQDTCRQGHPGIPGNPGHNGLPGRDGRDGAKGDKGDAGEPGRPGSPGKDGTSGEKGE RGADGKVEAKGIKGDQGSRGSPGKHGPKGLAGPMGEKGLRGETGPQGQKGNKGDVGPT GPEGPRGNIGPLGPTGLPGPMGPIGKPGPKGEAGPTGPQGEPGVRGIRGWKGDRGEKG KIGETLVLPKSAFTVGLTVLSKFPSSDMPIKFDKILYNEFNHYDTAAGKFTCHIAGVY YFTYHITVFSRNVQVSLVKNGVKILHTKDAYMSSEDQASGGIVLQLKLGDEMWLQVTG GERFNGLFADEDDDTTFTGFLLFSSPLE

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 11B.

Table 11B. Comparison of NOV11a against NOV11b through NOV11o.		
Protein Sequence	**NOV11a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV11b	25..231 72..314	183/243 (75%) 188/243 (77%)
NOV11c	101..231 1..131	131/131 (100%) 131/131 (100%)
NOV11d	25..230 91..332	183/242 (75%) 187/242 (76%)
NOV11e	1..230 1..287	209/287 (72%) 213/287 (73%)
NOV11f	100..231 2..133	131/132 (99%) 131/132 (99%)
NOV11g	100..231 2..133	131/132 (99%) 132/132 (99%)
NOV11h	100..231 2..133	132/132 (100%) 132/132 (100%)
NOV11i	25..231 74..316	182/243 (74%) 187/243 (76%)
NOV11j	25..231 74..316	182/243 (74%) 187/243 (76%)
NOV11k	25..231 74..316	183/243 (75%) 188/243 (77%)
NOV11l	25..231 74..316	182/243 (74%) 187/243 (76%)
NOV11m	25..231 74..316	184/243 (75%) 189/243 (77%)
NOV11n	25..231 74..316	183/243 (75%) 188/243 (77%)
NOV11o	25..231 74..316	183/243 (75%) 188/243 (77%)

Further analysis of the NOV11a protein yielded the following properties shown in Table 11C.

Table 11C. Protein Sequence Properties NOV11a	
PSort analysis:	0.6400 probability located in microbody (peroxisome); 0.5057 probability located in mitochondrial matrix space; 0.2277 probability located in mitochondrial inner membrane; 0.2277 probability located in mitochondrial intermembrane space
SignalP analysis:	Cleavage site between residues 20 and 21

A search of the NOV11a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 11D.

Table 11D. Geneseq Results for NOV11a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB27230	Human EXMAD-8 SEQ ID NO: 8 - Homo sapiens, 306 aa. [WO200068380-A2, 16-NOV-2000]	24..230 99..305	157/234 (67%) 164/234 (69%)	2e-79
AAW09108	Human adipocyte complement related protein Acrp30 - Homo sapiens, 244 aa. [WO9639429-A2, 12-DEC-1996]	23..228 36..240	110/207 (53%) 140/207 (67%)	1e-58
AAG80254	Human APM1 protein - Homo sapiens, 244 aa. [WO200132868-A1, 10-MAY-2001]	23..228 36..240	110/207 (53%) 140/207 (67%)	3e-58
AAB50373	Human adipocyte complement related protein ACRP30 - Homo sapiens, 244 aa. [WO200073448-A1, 07-DEC-2000]	23..228 36..240	110/207 (53%) 140/207 (67%)	3e-58
AAB49598	Human ACRP30 protein - Homo sapiens, 244 aa. [WO200073446-A2, 07-DEC-2000]	23..228 36..240	110/207 (53%) 140/207 (67%)	3e-58

5

In a BLAST search of public sequence databases, the NOV11a protein was found to have homology to the proteins shown in the BLASTP data in Table 11E.

Table 11E. Public BLASTP Results for NOV11a				
Protein Accession Number	Protein/Organism/Length	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q15848	Adiponectin precursor (30 kDa adipocyte complement-related protein) (ACRP30) (Adipose most abundant gene transcript 1) (apM-1) (Gelatin- binding protein) - Homo sapiens (Human), 244 aa.	23..228 36..240	110/207 (53%) 140/207 (67%)	8e-58
Q95JD7	ADIPONECTIN - Macaca mulatta (Rhesus macaque), 243 aa.	23..228 35..239	110/207 (53%) 141/207 (67%)	1e-57
Q95MQ4	ADIPOSE TISSUE-SPECIFIC PROTEIN ADIPO Q - Bos taurus (Bovine), 240 aa.	6..228 9..235	119/233 (51%) 151/233 (64%)	2e-56
Q60994	Adiponectin precursor (30 kDa adipocyte complement-related protein) (ACRP30) (Adipocyte specific protein AdipoQ) - Mus musculus (Mouse), 247 aa.	22..228 38..243	113/211 (53%) 142/211 (66%)	2e-56
Q95J95	ADIPONECTIN - Canis familiaris (Dog), 194 aa (fragment).	29..205 19..194	95/178 (53%) 121/178 (67%)	3e-48

PFam analysis predicts that the NOV11a protein contains the domains shown in
5   the Table 11F.

Table 11F. Domain Analysis of NOV11a			
Pfam Domain	NOV11a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Collagen	29..88	29/60 (48%) 47/60 (78%)	2.3e-12
C1q	101..227	58/141 (41%) 99/141 (70%)	1.2e-42

EXAMPLE 12.

The NOV12 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 12A.

Table 12A. NOV12 Sequence Analysis		
	SEQ ID NO: 67	851 bp
NOV12a, CG95794-01 DNA Sequence	TGGCCAGAATGTCAAGACAAGTCCGTTGGTATAAATGGCTGTGGAAGAAAAGTGTCAT CCTATTTCCAAGTCCAGGTGAAGTGAGCAACCACGAAGACCTTCATCTTCCTGCTCTC CTGGGAGCTGCTGCCACTTTCCCCACTAGTGATGATGATGACAAGATCGTTGGGGGGC TACACCTGTCAGACGAATGCTGTCCCCTATCAGGGTCCCTGAATGCTGGCTATCACTT CTGTGGTGGCTCCCTCATCAATAACAAATGGGAGGTGTCCACGGCTCACTGCTATAAG TCCCGAATCCATGTGCATCTTGGAGAATACAACATTAAGGTCTATGAAGGCAATGAAC AATTCATAAATGCAGCCAAGATTATTTGCCACCCCAAGTATAACTCAGCCACCATTGA TAATGACATCATGCTGATTAAGCTGAGCTCAGTCGCCACCATCAACTCTCAAGTGGCC ACCATCTCTCTGCCAAGATCCTGTGCAGCGGCTGGTACTCAGTGCCTCATCTCTGGCT GGGACAACACCCTGAGCAGTGGCTCCAACTACCCTGATCTCCTGCAGTGTCTGAAGGC TCCCATTCTCTCTAACACTGCTTGCCGCACAGCCTACCCAGGCAAGATTACTACAAAC ATGATATGTCTGGGATTCCTGGAGGGTGGAAAGGACTCTTGCCAGGGTGACTCTGGTG TTCCTGTGGTCTGCAACGGAGAACTCCAGGGCATTGTCTCCTGGGGCTATGGTTGTCC TCAGAAGAACAAACCTGGAGTCTACACTAAAGTTTGCAACTACGTGAAATGGATTCAG CAGACCATTGCTGCCAACTAAACACCTTTATCTCTTCAT
	ORF Start: ATG at 9	ORF Stop: TAA at 831
	SEQ ID NO: 68	274 aa MW at 29764.6kD
NOV12a, CG95794-01 Protein Sequence	MSRQVRWYKWLWKKSVILFPSPGEVSNHEDLHLPALLGAAATFPTSDDDDKIVGGLHL SDECCPLSGSLNAGYHFCGGSLINNKWEVSTAHCYKSRIHVHLGEYNIKVYEGNEQFI NAAKIICHPKYNSATIDNDIMLIKLSSVATINSQVATISLPRSCAAAGTQCLISGWDN TLSSGSNYPDLLQCLKAPILSNTACRTAYPGKITTNMICLGFLEGGKDSCQGDSGVPV VCNGELQGIVSWGYGCPQKNKPGVYTKVCNYVKWIQQTIAAN

Further analysis of the NOV12a protein yielded the following properties shown in

Table 12B.

Table 12B. Protein Sequence Properties NOV12a	
PSort analysis:	0.5729 probability located in mitochondrial matrix space; 0.2867 probability located in mitochondrial inner membrane; 0.2867 probability located in mitochondrial intermembrane space; 0.2867 probability located in mitochondrial outer membrane
SignalP analysis:	Cleavage site between residues 24 and 25

A search of the NOV12a protein against the Geneseq database, a proprietary

database that contains sequences published in patents and patent publication, yielded

several homologous proteins shown in Table 12C.

Table 12C. Geneseq Results for NOV12a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV12a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAW08475	Porcine trypsinogen - Sus scrofa, 247 aa. [WO9700316-A1, 03-JAN-1997]	33..274 7..247	187/242 (77%) 209/242 (86%)	e-109
AAY78974	Canine cationic trypsinogen amino acid sequence - Canis familiaris, 247 aa. [WO200009739-A1, 24-FEB-2000]	35..274 8..246	182/240 (75%) 199/240 (82%)	e-106
AAY78975	Canine anionic trypsinogen amino acid sequence - Canis familiaris, 246 aa. [WO200009739-A1, 24-FEB-2000]	35..274 8..246	179/240 (74%) 199/240 (82%)	e-104
AAB35701	Human trypsin hL amino acid sequence - Homo sapiens, 247 aa. [JP2000253887-A, 19-SEP-2000]	35..274 8..247	171/240 (71%) 198/240 (82%)	e-104
AAB80953	Bovine met-phe-trypsinogen - Bos sp, 231 aa. [WO200119970-A2, 22-MAR-2001]	47..274 4..231	166/228 (72%) 194/228 (84%)	6e-99

In a BLAST search of public sequence databases, the NOV12a protein was found

to have homology to the proteins shown in the BLASTP data in Table 12D.

Table 12D. Public BLASTP Results for NOV12a				
Protein Accession Number	Protein/Organism/Length	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P08426	Trypsin III, cationic precursor (EC 3.4.21.4) (Pretrypsinogen III) - Rattus norvegicus (Rat), 247 aa.	35..274 8..247	176/240 (73%) 202/240 (83%)	e-106
P00761	Trypsin precursor (EC 3.4.21.4) - Sus scrofa (Pig), 231 aa.	43..274 1..231	180/232 (77%) 202/232 (86%)	e-106
P06872	Trypsin, anionic precursor (EC 3.4.21.4) - Canis familiaris (Dog), 247 aa.	35..274 8..246	182/240 (75%) 199/240 (82%)	e-106
P06871	Trypsin, cationic precursor (EC 3.4.21.4) - Canis familiaris (Dog), 246 aa.	35..274 8..246	179/240 (74%) 199/240 (82%)	e-103
Q9CPN9	2210010C04RIK PROTEIN (TRYPSINOGEN 7) - Mus musculus (Mouse), 247 aa.	35..274 8..247	171/240 (71%) 198/240 (82%)	e-103

PFam analysis predicts that the NOV12a protein contains the domains shown in

the Table 12E.

Table 12E. Domain Analysis of NOV12a			
Pfam Domain	NOV12a Match Region	Identities/ Similarities for the Matched Region	Expect Value
trypsin	52..267	108/262 (41%) 185/262 (71%)	7.7e-89

5

EXAMPLE 13.

The NOV13 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 13A.

Table 13A. NOV13 Sequence Analysis	
SEQ ID NO: 69	818 bp

NOV13a, CG95804-01 DNA Sequence	<u>TGGACTCCTGTTACC</u>ATGAGGTTCCTGATCCTGTTCCTAGCCCTGTCCCTAGGAGGGA TTGATGCTGCACCTCCTGTCCAGTCTCGAATTGTTGGAGGATTTAACTGTGAGAAGAA TTCCCAGCCCTGGCAAGTGGCTGTGTACCGCTTCACCAAATATCAATGTGGGGGTATC CTGCTGAACGCCAACTGGGTTCTCACAGCTGCCCACTGCCATAATGACAAGTACCAGG TGTGGCTGGGCAAAAACAACTTTTTGGAGGATGAACCCTCTGCCCAACACCGGCTTGT CAGCAAAGCCATCCCTCACCCTGACTTCAACATGAGCCTCCTGAATGAGCACACCCCA CAACCTGAGGATGACTACAGCAATGACCTGATGCTGCTCCGCCTCAAAAAGCCTGCTG ACATCACAGATGTTGTGAAGCCCATCGACCTGCCCACTGAGGAGCCCAAGCTGGGGAG CACATGCCTAGCCTCAGGCTGGGGCAGCATTACACCCGTCAAATATGAATACCCAGAT GAGCTCCAGTGTGTGAACCTCAAGCTCCTGCCTAATGAGGACTGTGCCAAAGCCCACA TAGAGAAGGTGACAGATGACATGTTGTGTGCAGGAGATATGGATGGAGGCAAAGACAC TTGTGCGGGTGACTCAGGAGGCCCACTGATCTGTGATGGTGTTCTCCAAGGTATCACA TCATGGGGCCCTAAGCCTTGCGGTAAACCCAATGTGCCGGGTATCTACACCAGAGTTT TAAATTTCAACACCTGGATAAGAGAAACTATGGCTGAAAATGACT<u>GAGTATCACATTG</u> <u>TCCCAT</u>
	ORF Start: ATG at 16	ORF Stop: TGA at 799
	SEQ ID NO: 70	261 aa	MW at 28815.6kD
NOV13a, CG95804-01 Protein Sequence	MRFLILFLALSLGGIDAAPPVQSRIVGGFNCEKNSQPWQVAVYRFTKYQCGGILLNAN WVLTAAHCHNDKYQVWLGKNNFLEDEPSAQHRLVSKAIPHPDFNMSLLNEHTPQPEDD YSNDLMLLRLKKPADITDVVKPIDLPTEEPKLGSTCLASGWGSITPVKYEYPDELQCV NLKLLPNEDCAKAHIEKVTDDMLCAGDMDGGKDTCAGDSGGPLICDGVLQGITSWGPK PCGKPNVPGIYTRVLNFNTWIRETMAEND

Further analysis of the NOV13a protein yielded the following properties shown in Table 13B.

Table 13B. Protein Sequence Properties NOV13a	
PSort analysis:	0.6281 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Cleavage site between residues 18 and 19

A search of the NOV13a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 13C.

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB21319	Human KLK2 - Homo sapiens, 262 aa. [WO200053776-A2, 14-SEP-2000]	1..260 1..261	170/261 (65%) 214/261 (81%)	e-105
AAB54293	Human pancreatic cancer antigen protein sequence SEQ ID NO:745 - Homo sapiens, 267 aa. [WO200055320-A1, 21-SEP-2000]	1..260 6..266	169/261 (64%) 214/261 (81%)	e-105
AAW71005	Human prostate-associated kallikrein designated HPAK - Homo sapiens, 262 aa. [WO9832865-A1, 30-JUL-1998]	1..260 1..261	169/261 (64%) 214/261 (81%)	e-105
AAP95121	Kallikrein encoded by clone lambda HK65a - Homo sapiens, 262 aa. [EP297913-A, 04-JAN-1989]	1..260 1..261	169/261 (64%) 214/261 (81%)	e-105
AAP70568	Human kallikrein-like substance has hypotensive activity - Homo sapiens, 262 aa. [JP62126980-A, 09-JUN-1987]	1..260 1..261	169/261 (64%) 213/261 (80%)	e-104

5

In a BLAST search of public sequence databases, the NOV13a protein was found to have homology to the proteins shown in the BLASTP data in Table 13D.

Table 13D. Public BLASTP Results for NOV13a				
Protein Accession Number	**Protein/Organism/Length**	**NOV13a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
P15947	Glandular kallikrein K1 precursor (EC 3.4.21.35) (Tissue kallikrein) (mGK-6) (Renal kallikrein) (KAL-B) - Mus musculus (Mouse), 261 aa.	1..261 1..261	260/261 (99%) 260/261 (99%)	e-159
A25606	tissue kallikrein (EC 3.4.21.35) submandibular precursor - mouse, 261 aa.	1..261 1..261	259/261 (99%) 259/261 (99%)	e-158
P15945	Glandular kallikrein K5 precursor (EC 3.4.21.35) (Tissue kallikrein) (MGK-5) - Mus musculus (Mouse), 261 aa.	1..260 1..260	237/260 (91%) 247/260 (94%)	e-145
P00757	7S nerve growth factor alpha chain precursor (Alpha-NGF) - Mus musculus (Mouse), 256 aa.	1..260 1..255	214/260 (82%) 234/260 (89%)	e-129
P32824	Glandular kallikrein, renal precursor (EC 3.4.21.35) (Tissue kallikrein) - Praomys natalensis (African soft-furred rat) (Mastomys natalensis), 263 aa.	1..260 1..262	210/262 (80%) 233/262 (88%)	e-127

PFam analysis predicts that the NOV13a protein contains the domains shown in the Table 13E.

Table 13E. Domain Analysis of NOV13a			
Pfam Domain	**NOV13a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Trypsin	25..253	100/268 (37%) 199/268 (74%)	8.4e-104

5

EXAMPLE 14.

The NOV14 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 14A.

Table 14A. NOV14 Sequence Analysis

	SEQ ID NO: 71	2691 bp
NOV14a, CG95861-01 DNA Sequence	GCTTGCCCGTCGGTCGCTAGCTCGCTCGGTGCGCGTCGTCCCGCTCC**ATG**GCGCTCTT CGTGCGGCTGCTGGCTCTCGCCCTGGCTCTGGCCCTGGGCCCCGCCGCGACCCTGGCG GGTCCCGCCAAGTCGCCCTACCAGCTGGTGCTGCAGCACAGCAGGCTCCGGGGCCGCC AGCACGGCCCCAACGTGTGTGCTGTGCAGAAGGTTATTGGCACTAATAGGAAGTACTT CACCAACTGCAAGCAGTGGTACCAAAGGAAAATCTGTGGCAAATCAACAGTCATCAGC TACGAGTGCTGTCCTGGATATGAAAAGGTCCCTGGGGAGAAGGGCTGTCCAGCAGCCC TACCACTCTCAAACCTTTACGAGACCCTGGGAGTCGTTGGATCCACCACCACTCAGCT GTACACGGACCGCACGGAGAAGCTGAGGCCTGAGATGGAGGGGCCCGGCAGCTTCACC ATCTTCGCCCCTAGCAACGAGGCCTGGGCCTCCTTGCCAGCTGAAGTGCTGGACTCCC TGGTCAGCAATGTCAACATTGAGCTGCTCAATGCCCTCCGCTACCATATGGTGGGCAG GCGAGTCCTGACTGATGAGCTGAAACACGGCATGACCCTCACCTCTATGTACCAGAAT TCCAACATCCAGATCCACCACTATCCTAATGGGATTGTAACTGTGAACTGTGCCCGGC TCCTGAAAGCCGACCACCATGCAACCAACGGGGTGGTGCACCTCATCGATAAGGTCAT CTCCACCATCACCAACAACATCCAGCAGATCATTGAGATCGAGGACACCTTTGAGACC CTTCGGGCTGCTGTGGCTGCATCAGGGCTCAACACGATGCTTGAAGGTAACGGCCAGT ACACGCTTTTGGCCCCGACCAATGAGGCCTTCGAGAAGATCCCTAGTGAGACTTTGAA CCGTATCCTGGGCGACCCAGAAGCCCTGAGAGACCTGCTGAACAACCACATCTTGAAG TCAGCTATGTGTGCTGAAGCCATCGTTGCGGGGCTGTCTGTAGAGACCCTGGAGGGCA CGACACTGGAGGTGGGCTGCAGCGGGGACATGCTCACTATCAACGGGAAGGCGATCAT CTCCAATAAAGACATCCTAGCCACCAACGGGGTGATCCACTACATTGATGAGCTACTC ATCCCAGACTCAGCCAAGACACTATTTGAATTGGCTGCAGAGTCTGATGTGTCCACAG CCATTGACCTTTTCAGACAAGCCGGCCTCGGCAATCATCTCTCTGGAAGTGAGCGGTT GACCCTCCTGGCTCCCCTGAATTCTGTATTCAAAGATGGAACCCCTCCAATTGATGCC CATACAAGGAATTTGCTTCGGAACCACATAATTAAAGACCAGCTGGCCTCTAAGTATC TGTACCATGGACAGACCCTGGAAACTCTGGGCGGCAAAAAACTGAGAGTTTTTGTTTA TCGTAATAGCCTCTGCATTGAGAACAGCTGCATCGCGGCCCACGACAAGAGGGGGAGG TACGGGACCCTGTTCACGATGGACCGGGTGCTGACCCCCCCAATGGGGACTGTCATGG ATGTCCTGAAGGGAGACAATCGCTTTAGCATGCTGGTAGCTGCCATCCAGTCTGCAGG ACTGACGGAGACCCTCAACCGGGAAGGAGTCTACACAGTCTTTGCTCCCACAAATGAA GCCTTCCGAGCCCTGCCACCAAGAGAACGGAGCAGACTCTTGGGAGATGCCAAGGAAC TTGCCAACATCCTGAAATACCACATTGGTGATGAAATCCTGGTTAGCGGAGGCATCGG GGCCCTGGTGCGGCTAAAGTCTCTCCAAGGTGACAAGCTGGAAGTCAGCTTGAAAAAC AATGTGGTGAGTGTCAACAAGGAGCCTGTTGCCGAGCCTGACATCATGGCCACAAATG GCGTGGTCCATGTCATCACCAATGTTCTGCAGCCTCCAGCCAACAGACCTCAGGAAAG AGGGGATGAACTTGCAGACTCTGCGCTTGAGATCTTCAAACAAGCATCAGCGTTTTCC AGGGCTTCCCAGAGGTCTGTGCGACTAGCCCCTGTCTATCAAAAGTTATTAGAGAGGA TGAAGCAT**TAG**CTTGAAGCACTACAGGAGGAATGCACCACGGCAGC<u>TCTCCGCCAATT</u> <u>TCTCTCAGATTTCCACAGAGACTGTTTGAATGTTTTCAAAACCAAGTATCACACTTTA</u> <u>ATGTACATGGGCCGCACCATAATGAGATGTGAGCCTTGTGCATGTGGGGGAGGAGGGA</u> <u>GAGAGATGTACTTTTTAAATCATGTTCCCCCTAAACATGGCTGTTAACCCACTGCATG</u> <u>CAGAAACTTGGATGTCACTGCCTGACATTCACTTCCAGAGAGGACCTATCCCAAATGT</u> <u>GGAATTGACTGCCTATGCCAAGTCCCTGGAAAAGGAGCTTCAGTATTGTGGGGCTCAT</u> <u>AAAACATGAATCAAGCAATCCAGCCTCATGGGAAGTCCTGGCACAGTTTTTGTAAAGC</u> <u>CCTTGCACAGCTGGAGAAATGGCATCATTATAAGCTATGAGTTGAAATGTTCTGTCAA</u> <u>ATGTGTCTCACATCTACACGTGGCTTGGAGGCTTTTATGGGGCCCTGTCCAGGTAGAA</u> <u>AAGAAATGGTATGTAGAGCTTAGATTTCCCTATTGTGACAGAGCCATGGTGTGTTTGT</u> <u>AATAATAAAACCAAAGAAACATA</u>

	ORF Start: ATG at 48	ORF Stop: TAG at 2097	
	SEQ ID NO: 72	683 aa	MW at 74680.0kD
NOV14a, CG95861-01 Protein Sequence	MALFVRLLALALALALGPAATLAGPAKSPYQLVLQHSRLRGRQHGPNVCAVQKVIGTN RKYFTNCKQWYQRKICGKSTVISYECCPGYEKVPGEKGCPAALPLSNLYETLGVVGST TTQLYTDRTEKLRPEMEGPGSFTIFAPSNEAWASLPAEVLDSLVSNVNIELLNALRYH MVGRRVLTDELKHGMTLTSMYQNSNIQIHHYPNGIVTVNCARLLKADHHATNGVVHLI DKVISTITNNIQQIIEIEDTFETLRAAVAASGLNTMLEGNGQYTLLAPTNEAFEKIPS		

| | ETLNRILGDPEALRDLLNNHILKSAMCAEAIVAGLSVETLEGTTLEVGCSGDMLTING KAIISNKDILATNGVIHYIDELLIPDSAKTLFELAAESDVSTAIDLFRQAGLGNHLSG SERLTLLAPLNSVFKDGTPPIDAHTRNLLRNHIIKDQLASKYLYHGQTLETLGGKKLR VFVYRNSLCIENSCIAAHDKRGRYGTLFTMDRVLTPPMGTVMDVLKGDNRFSMLVAAI QSAGLTETLNREGVYTVFAPTNEAFRALPPRERSRLLGDAKELANILKYHIGDEILVS GGIGALVRLKSLQGDKLEVSLKNNVVSVNKEPVAEPDIMATNGVVHVITNVLQPPANR PQERGDELADSALEIFKQASAFSRASQRSVRLAPVYQKLLERMKH |

Further analysis of the NOV14a protein yielded the following properties shown in Table 14B.

Table 14B. Protein Sequence Properties NOV14a	
Psort analysis:	0.8200 probability located in endoplasmic reticulum (membrane); 0.1900 probability located in plasma membrane; 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Cleavage site between residues 24 and 25

A search of the NOV14a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 14C.

Table 14C. Geneseq Results for NOV14a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM24494	Colon tumour related amino acid sequence SEQ ID NO:122 - Homo sapiens, 683 aa. [WO200149716-A2, 12-JUL-2001]	1..683 1..683	683/683 (100%) 683/683 (100%)	0.0
AAB11897	Human colon tumour polypeptide, SEQ ID NO:122 - Homo sapiens, 683 aa. [WO200037643-A2, 29-JUN-2000]	1..683 1..683	683/683 (100%) 683/683 (100%)	0.0
AAR80573	Human beta-IG-H3 (transforming growth factor-beta induced gene-h3) - Homo sapiens, 683 aa. [US5444164-A, 22-AUG-1995]	1..683 1..683	683/683 (100%) 683/683 (100%)	0.0
AAR40386	betaIG-H3 protein - Homo sapiens, 683 aa. [EP555989-A, 18-AUG-1993]	1..683 1..683	683/683 (100%) 683/683 (100%)	0.0
AAR74302	TCI protein - Homo sapiens, 777 aa. [WO9511923-A, 04-MAY-1995]	1..682 1..683	327/687 (47%) 453/687 (65%)	0.0

In a BLAST search of public sequence databases, the NOV14a protein was found to have homology to the proteins shown in the BLASTP data in Table 14D.

Table 14D. Public BLASTP Results for NOV14a				
Protein Accession Number	Protein/Organism/Length	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q15582	Transforming growth factor-beta induced protein IG-H3 precursor (Beta IG-H3) (Kerato-epithelin) (RGD-containing collagen associated protein) (RGD-CAP) - Homo sapiens (Human), 683 aa.	1..683 1..683	683/683 (100%) 683/683 (100%)	0.0
O11780	Transforming growth factor-beta induced protein IG-H3 precursor (Beta IG-H3) (Kerato-epithelin) (RGD-containing collagen associated protein) (RGD-CAP) - Sus scrofa (Pig), 683 aa.	1..683 1..683	633/683 (92%) 664/683 (96%)	0.0
Q95215	Transforming growth factor-beta induced protein IG-H3 precursor (Beta IG-H3) (Kerato-epithelin) (RGD-containing collagen associated protein) (RGD-CAP) - Oryctolagus cuniculus (Rabbit), 683 aa.	1..683 1..683	630/683 (92%) 656/683 (95%)	0.0
P82198	TRANSFORMING GROWTH FACTOR-BETA INDUCED PROTEIN IG-H3 PRECURSOR (BETA IG-H3) (KERATO-EPITHELIN) (RGD-CAP) (P68 BETA IG-H3) - Mus musculus (Mouse), 683 aa.	1..683 1..683	619/683 (90%) 652/683 (94%)	0.0
O42390	TRANSFORMING GROWTH FACTOR-BETA INDUCED PROTEIN IG-H3 PRECURSOR (BETA IG-H3) (KERATO-EPITHELIN) (RGD-CAP) - Gallus gallus (Chicken), 680 aa.	12..682 9..674	529/671 (78%) 607/671 (89%)	0.0

PFam analysis predicts that the NOV14a protein contains the domains shown in
5   the Table 14E.

160

Table 14E. Domain Analysis of NOV14a			
Pfam Domain	**NOV14a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Fasciclin	113..238	43/150 (29%) 84/150 (56%)	1.1e-10
Fasciclin	240..373	44/149 (30%) 94/149 (63%)	1.5e-26
Fasciclin	376..500	37/150 (25%) 81/150 (54%)	0.00032
Fasciclin	502..634	71/149 (48%) 128/149 (86%)	2.7e-67

EXAMPLE 15.

The NOV15 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 15A.

Table 15A. NOV15 Sequence Analysis		
	SEQ ID NO: 73	8279 bp
NOV15a, CG96412-01 DNA Sequence	AGTCGGGCCCCTCGGGGCCGCGTGGCCAATCAGATCCCCCCTGAGATCCTGAAGAACC CTCAGCTGCAGGCAGCAATCCGGGTCCTGCCTTCCAACTACAACTTTGAGATCCCCAA GACCATCTGGAGGATCCAACAAGCCCAGGCCAAGAAGGTGGCCTTGCAA**ATG**CCGGAA GGCCTCCTCCTCTTTGCCTGTACCATTGTGGATATCTTGGAAAGGTTCACGGAGGCCG AAGTGATGGTGATGGGTGACGTGACCTACGGGGCTTGCTGTGTGGATGACTTCACAGC GAGGGCCCTGGGAGCTGACTTCTTGGTGCACTACGGCCACAGTTGCCTGAGTATGGTA GATCTTTCCTTTGGATTTGGTTGCCTTGGCAACGGTGCTCTGTCCCAGATGCAGGTGT TTGAAAGGCTGTTGGTTGTAGAGCAGGCTGGGCCCCGGCCGGTTCCCATGGACACCTC GGCCCAAGACTTCCGGGTGCTGTACGTCTTTGTGGACATCCGGATAGACACTACACAC CTCCTGGACTCTCTCCGCCTCACCTTTCCCCCAGCCACTGCCCTTGCCCTGGTCAGCA CCATTCAGTTTGTGTCATTCCTACAGGCAGCCGCCCAGGAGCTGAAAGCCGAGTATCG TGTGAGTGTCCCACAGTGCAAGCCCCTGTCCCCTGGAGAGATCCTGGGCTGCACATCC CCCCGACTGTCCAAAGAGGTGGAGGCCGTTGTATCCGCCGCGGCATTAGATTCTTGTA GGAGCTCAAACCCTATCGTGACCTGTACATGCGAGGGATCTAGGTTGCATGATCTTTA TGAGATCCTAATGCCTGATGATCTGAGGTATCTTGGAGATGGCCGCTTCCATCTGGAG TCTGTCATGATTGCCAACCCCAATGTCCCCGCTTACCGGTATGGGCTGGGCCGGGCTG GGCTGACCAGCTGGTATGACCCATATAGCAAAGTCCTATCCAGAGAACACTATGACCA CCAGCGCATGCAGGCTGCTCGCCAAGAAGCCATAGCCACTGCCCGCTCAGCTAAGTCC TGGGGCCTTATTCTGGGCACTTTGGGCCGCCAGGGCAGTCCTAAGATCCTGGAGCACC TGGAATCTCGACTCCGAGCCTTGGGCCTTTCCTTTGTGAGGCTGCTGCTCTCTGAGAT CTTCCCCAGCAAGCTTAGCCTACTTCCTGAGGTGGATGTGTGGGTGCAGGTGGCATGT CCACGTCTCTCCATTGACTGGGGCACAGCCTTCCCCAAGCCGCTGCTGACACCCTATG AGGTAACACCAAGCTCTGGGAGAGAGTGGGCTTTGGACGTGGTTCTCAAAGGCGGCCG TGGCTCTGAGGGACATTTCCTGGCAGCAGCCCTACCCGATGGACTTCTACGCTGGCAG CTCCTTGGGGCCCTGGACGGTGAACCACGGCCAGGACCGCCGTCCCCACGCCCCGGGC CGGCCCGCGCGGGGGAAGGAGGGGGTCCGCGCGTCCCCCTTCGGCCGTGGCTTGCGAGG ACTGCAGCTGCAGGGACGAGAAGGTGGCGCCGCTGGCTCCTTGACGCGCTCCCGGGCC TCAGACCGCTTCCGGTGCTTCCGTCGCTCCTTGCCGGGCATAATGGCCGCGCAGCGAC CCCTGCGGGTCCTGTGCCTGGCGGGCTTCCGGCAGAGCGAGCGGGGCTTCCGTGAGAA GACCGGGGCGCTGAGGAAGGCGCTGCGGGGTCGCGCCGAGCTCGTGTGCCTCAGCGGC CCGCACCCGGTCCCCGACCCCCCGGGCCCCGAGGGCGCCAGATCAGACTTCGGGTCCT GCCCTCCGGAGGAGCAGCCTCGAGGCTGGTGGTTTTCAGAGCAGGAGGCCGACGTTTT CTCCGCATTGGAAGAGCCCGCCGTCTGCAGGGGCCTGGAGGAATCACTGGGGATGGTG GCACAGGCACTGAACAGGCTGGGGCCTTTTGACGGCCTTCTTGGTTTCAGCCAAGGGG CTGCGCTAGCAGCCCTTGTGTGTGCCCTGGGCCAGGCAGGCGATCCCCGCTTCCCCTT GCCACGGTTTATCCTCTTGGTGTCTGGTTTCTGTCCCCGGGGCATTGGGTTCAAGGAA TCCATCCTGCAAAGGCCCTTGTCATTGCCTTCGCTCCATGTTTTTGGGGACACTGACA AAGTCATCCCCTCTCAGGAGAGTGTGCAACTGGCCAGCCAATTTCCCGGAGCCATCAC CCTCACCCACTCTGGTGGCCACTTCATTCCAGCAGCTGCACCCCAGCGTCAGGCCTAC CTCAAGTTCTTGGACCAGTTTGCAGAG**TGA**AAGATCAAGAAATGTCTCTGCTCCTACA TCCAGCTCCTCTAGGGGCAGCCTCCGTCATCCATGCCCTCCCAGGACCCTCCACTCAC TGCTGTGAGTGCGCCTCACCAGAACCAGTTAAGAGACAACTATCAATTCTTGAGACCC AAATTATAAGGGCCCTGCCCTGTACTGAAGAAAAGGGGAGCACAAGGCCTTAATGGAC ATTGACTTGTGAAAACGCAAACATGAATATGGTTGGAGAGCCCTGGATTAGGAGGGTG ACATGGGAAGGCAGAGGCTGGCACCATGGTGACTGCCACATAATAAAGTGGTGATTT GGATTTTGAGCATCTTTTTCCTGGTACACACAGAAAAACATTTTATAATGGAAGTCGG TTTCTTGGCCATCTACATAGTTTTCTGGGCAGCGCCAAGCAGGGAGGTGTCTCCAGTT GTGAGTCCTCGGACAGGCTGCTGCATGGGTGCACATACTCACGTTATTGGTGGAAGTT TTAAGTCCCAAACTGAAGGGAAGGAGGCCTAGGTGGCACAATCAGGGAGGAATATACA TCTGAGAAGTTTTGGGAAGACATCACCTGGCAAGGCTGCCTGAACCACAGTAATTTAG

```
TCTTCCTCTATCCAGATCACTGAGAGTTGTGTGCTGGTCTTCTCTCAAACCCTGGATG
TGTTACCTGCTTCTCCCTAAGTGCTCTAACCTCGCTGAACTATGAGCCTGAAGGGTGG
GTCGATGCCAGACCTTAGGTGGAGGCCAAAGACAGTACAGAATAAACAGCTTCTTCCT
AAATTGACCCCATCTTGAGGGTTTCTGAGATTGATGCCATTCTTTAGTGACTACAGCC
AAGGCCTAAGCAATCCAGCTGGTTTTCCCCTTGGGGCGTGTAGTTGTTCTTGGATACT
TTGAGGATTGCTCACTCTTTTTTTTTTTTTTTTTGAGACAGGCTCACTCTGTCACCCA
GGCTGGAGAGCAGTGGCCAAGATCCTGGCTCACTGCAACCTCAGCCTCCTGGGTTCAA
GCAGTTCTGTCTCAGCCTCCTGAGTAGGTGGCGTTACAGGCATGCGCCACCACGCCTG
GCTAATTTCTATATTTTTTTGTAGAGATAGGGTTTTGCCACGTTGCCCAGGCTGGTCT
CAAAACTCTTTAGCTCAAACAATCCACCTGCCTCAGCCTCCCAAAGTGCTAGGATTAT
AGGCCTGAGCCACTGCACCTACCTAGGATTGCCCACTCTTTTTTTGTTTTCTTTACTT
TTAAAGACAAGGTCTGACTCTATCACCCAGGCTGGAGTGCAGTGGCACAATCTTGACA
GCCTCCACCTCCCGGGCTCAAGCCATCCATCCTCCCACCTCAGCCTCCCAAGCAGCTG
GGGCTGCAGGCGCAAATCACCATGCCTGGCTAATTTTTGTATTTGTTGTAGAGAAGGG
GTCTCACCATGTTGTGCGGATGGTCTCCAACTTGCGAGCTCAAGCAATCCACCTGTCC
TGGCCTCCCAAAGTGCTGGAATTAACAGGCGTGAGCCACCTCACCTGGCCAATTGCCC
ACTCTTAATGAGTTCCCTCCCATTTCCAGTGCCACCTGACACCCCCACCCCCTGACA
CACACACACACTCTTGTTTTCTGACTTCCTGCCTCTGCAAGACAAGAAGCCTGTTC
TTTTTCTGGAAGATCACACCAACTCCTGTTTAGCCCTCAGACATAGTTGCCAGGGAGC
TAAGCAACTGAGCAGGAATTATGGTCAGGGACACAGCACAAATACTCCAGCCCATTTT
TCTGACAACCAGACTAATATCCCTTACATTACCGAAAGTTCTGGGTTCCTCACTACAG
AGATTAGCTGTCCAAAATTAAAAAAAACAAAACAAAACAGAATTATCCGCCCAGAGAC
CTAGTTACTGGGTGTGGAAAGTGTGTGTGGGGAGGGGTGGGAGGTGGTTCTCCCTTGC
TTCAGTTGGCTTCTGCCTTGTTTTAAGGAATCTTCTCCTTTAGCAAAAGGAGGAGACT
TTGGAATGGATTGATTACACAGACTGTGGTATCTGACCCATTGAATTTTAGAAAAAAT
TCTGATTTAAAGTTTCAGAAATTGCACAGAAAATTTCAGATTTCTGGCTTCTCTGGAA
ATGATAGATTTGCAGTTCAGGGCTCCCATACCCTCATGGTAATCATAGGCTGCCCCCT
TTAGCACGGTCCCAAGAGGTGAAAACGACTTGTACACCCCAGCTGCTTGGGATTATTG
AGGACGAAAGTAGAGAGAAGGGGAGAAATATTGGGGAGTGAATTTTGAGCCAAGGCTT
AAAATTAAAAGTGGGGGAGGTAGAGCCCAGCAGTAGTAGGTGGAGGAGAAGGGCTCCT
GGCCGGGGGTGAGGATTCCTCCTGAGAACCATAGTGTCCAAGCTAGAGGGAAACATGG
GGTCCACTAGTTCTCGACAAGTGGAACAGGTCACTTCCCATCATGCCGTCCAGGAGCA
AGAGAGACTCTTAGGGGGTTGTGCCCCTCCCCACCCCCAATACGCAGATTCTGTGTAC
CATTTACATCCAGAACATTGGTTAAAACCTGAATTCTGGCCCTGCGTGGTGGCTCATG
CCTGTAATCCCAGCACTTTGGGAGGCGGAGGCGGGCAGATCACCTGAGGTTGGGAATT
TGAGGCCAGCCAGACCAACAGAGAGAAACTCTGTCTCTACTAAAAATACAAAATTAGC
CAACCATGGTGGCGCATGCCTGTAATCCCAGCTACTTGGAAGGCTGAAGCAGGAGAAT
TGCTTGAACCTGGGAGGCAAAGGTTGCGGTAAGCCGAGATCACGCTATTGCACTCCAA
CCTGGGCAACGAGAGCGAAACTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAGTAGCTGGGCG
TGGTGGCATGCACCTGTAATCCTAGCTACTTGGTAGGCTGAGGCAGGAGAATTGCTTG
AACCCAGGAGGCGGAGGTTGCAGTGAGCCGAAATTGTGCCACTGCACTCCAGCCTGGG
TGACAGAGTGAGACTCCATCTCAAAAAACCAACTTGAATTCTGGGGCCAACTCCAGAT
GTACTGAGTCAGAATCACTGGGACCCAGGAATCTGCATTTTGGCAAATTGCCCCCATC
ATTCCCAAGCAGGGATTTGAAAAGCCACTGCCTTAGGGGTTGAATGAAAGCCTCGGGT
GAGGGTGTGTTATCTGAAAAACCCCTAGTGTGACTCTCTTGCAACCCACCAGGGAGGG
GAGCAAGCTCCCCAGAACAATCACTGCTTAAACATTTTCAATTTATAGCAAAGGAAAA
TTAGGTCTGAAGGATAGAAATAGGAATAGCATTTATTTATTTACTTACTTATTTGAGA
CGGAGTCTCGCTCTGTCACCCAGGCTGGAGTGCAGTGGTACAGTCTCCGCTCTCTGCG
ACCTCCGCCTCCTGAGTTCGAGCGATCCGCCTGCCTCAGCCTCCAGAGTAGCTGGGAT
TACAGGTGTGCATCACCACTCCCAGCTAATTTTTGCATTTTCAGTAGAGACAGGGTTT
CGCCATGTTGGCCAGGCTGGTCTTGAACTCCTGATCTCAAGTGATTCGCCCACCTCAG
CCTCCCAAAGTGTTGGGATTACAGGTGTGAGCCACTGCACCCGGCCAGAAGAGCATTT
ATTGAGTAATAACTGTATACCAGACATTGTAATAAGCAAGAGCTTTAATGCGGCCTAA
TTTAGGGCCAGGTGCGATTGCTCATGCCTATAATCCCAGCACTTTGGGAGGCTAAGGC
ATGAGGATCAATTGAGCCCAGGAGTTCAGAACCAGCCTGGGAAACATAGTGAGACCCC
CTCTCTACAAATAATTTAAAAGTTAGCTGGGTGTGGTGGCATGCACCTGTGGTCCCAG
CTACTCGGGAGGCTCAGGTGGGAGGATCACTTGAGCCCAGGAGTTTTGAGGCTGCAGT
AAGCTGTGATTGTGCCACTGTATTCCAGCCTGGGAGACAGAGCTAGACCCTGTCTCCA
AAAAAAAAGAAAAAAAAACAAACTACAAAAATTGGCTGGGGGTGGTGGCACACACCTG
```

	TAGTCCTAACTACTCAGGAGGCTGAGACAGGAGAATTGCTTGAACCTGGGAGGCGGAC GTTGCAGTGAGCCGAGATCATGCTACTGCACTTCAGCCTGGGCAACAGAGCAAGACTC CATCTCAGAAGAAAAAAACAAATAAAATTATATATATATAACACATATATATTTTATA TATATTAGGCAATAAATATATATAGCTTAAACCTTACAACTTCATGTTTAAACTCACA ACTTCATGAGGTAGGTACCATTATTATTATTATTATTTGAGACGGAGTGTCACTCAGC CACCCAGGCTGGAGTTCAGTGGCACGATCACGGCTCACTGCACCTCCGTCTCCTGGGT TCAAGTGATTCTAGTGCCTCAGCCTCCCAAGCAGCTGGAATTACTGGCAAGCACCACC ATGGCTGGCTAATTTTTTGTGTTATTAGTAGAGACGGAGTCTGACCATATTGGCCAGG CTGGATTATTATTAAGACAGGGTCTACCTCTGTCACCAGGTGCAGGATCATGGCTCAC TGCAGCCTCGACCTCTTCGGCTCAAGTGACCCTCCCGCTTCAGCCTCCTGAGTAGCTT AGCTGAGACTACAGGTGCCACCTATCACACTGGCTTTTTTTTTTTAAGACATGGGGT CTCACTATGTTGCCCAGGCTGGTCTCGAATTTCTGGGCTCAAGCAGTCCTCCAGCCTC TGCCCCGCCCCGCAAAATGGTGTGATTACAGGCATGAGCCATGTCTCCCAGCCCTAAA TAATGATTTTTTTTTTTTGAGACAGGGTCTCACTCTGTCATCCAGGCTGGAGTACAGT GGTGCAATCACAGCTCACTGCAGCCTCAACCTCCCAGCCTCAAGCAATCCTCCCACCT CAACTTCCCAAGTAGCTGGGACTACAGGTGTGCACAACTAATTTTCTTTCTTTTCTTT CCTTTCTTCTTTCCTTTCTTTCTTTTTTGTAGAGATGGGAGTCGCATTGTATTGCCCA GGCTGGTCTCGAACTCCTGGGCTCAAGAGATCCTCCCACCTATGCCTCCCAAAGTGCT GGGATTACAGGCGTGAGCCACTGTGCCCTGCCCTAAATGATTTCTAAGATGCTAAACA ACCCTGGCGTTCGGTCAAACCTCCGGTTCTTTCCCCCCCTACCACATTCGTGACCTTT TCTCAAATAATCCCAACCATTCTGTTTTTCCACCTCTGATTCCACCTTGGATTCCTAG AGCTCTAGAGAGAGAACCCCCTTAGCAAGAACCTCTCACAGCCCCAGGCCCAGAGTGA ACCCGGAAGGCGGAGGTTGCAGTGAGCCGAGATAGTGCCACTGCACTCCAGCCTGGGC GACAGAGCGAGACTGTCTCAGAAGAAAAAAAATAAAGTATATGTATGTATAACATAT ATATAATATAGATTATATATATATTAGCCAATAAATATATATAGCCTAAAGCTCAAGG TGTACCCCCCACATCCCGTTCATTCTGCCTAAGTGCACAGAGCTGGGGTCCCCTCCCA ACACCACCCCTTCACAAACATGTAGTAGCCTAACAGAATACTTAAGATCAAAATAACA TTCTGCTGTTTTTGGATTTTTACTTACATGATCCTATTTAATCCTCACAGTGCACCTG CCGAGGAGAAGAGATTATTATCCCGTTTTACAGAAAGCTCAGAAACAGAATATGGCCC CTTAACCACCAGTTGCTGATAATGCCAGGTTGGACCTCTGGCTTCCCAGAGCACTACA CCCAGCAGTCCACTGCAGTCCACCCACTGACATCCCCTGGAACCCAGAAGCAAGTGCG TTTTAATCTGCAAGACCTCGGGGCCCTGGGGAGGTGGATGGCTAGCATGTGGGTGTT GATTAACTGGGAAACCGGCACGAGTGTCTTAGAAGTACTTCACAAAGGAGCCGGGTGG GGAGTGAAGGAGGGGGTGGGGCGTGAGACGTTAAGAAAAATTG
	ORF Start: ATG at 166 **ORF Stop: TGA at 2290**
	SEQ ID NO: 74 708 aa MW at 76785.2kD
NOV15a, CG96412-01 Protein Sequence	MPEGLLLFACTIVDILERFTEAEVMVMGDVTYGACCVDDFTARALGADFLVHYGHSCL SMVDLSFGFGCLGNGALSQMQVFERLLVVEQAGPRPVPMDTSAQDFRVLYVFVDIRID TTHLLDSLRLTFPPATALALVSTIQFVSFLQAAAQELKAEYRVSVPQCKPLSPGEILG CTSPRLSKEVEAVVSAAALDSCRSSNPIVTCTCEGSRLHDLYEILMPDDLRYLGDGRF HLESVMIANPNVPAYRYGLGRAGLTSWYDPYSKVLSREHYDHQRMQAARQEAIATARS AKSWGLILGTLGRQGSPKILEHLESRLRALGLSFVRLLLSEIFPSKLSLLPEVDVWVQ VACPRLSIDWGTAFPKPLLTPYEVTPSSGREWALDVVLKGGRGSEGHFLAAALPDGLL RWQLLGALDGEPRPGPPSPRPGPARAGEGGVRASPFGRGLRGLQLQGREGGAAGSLTR SRASDRFRCFRRSLPGIMAAQRPLRVLCLAGFRQSERGFREKTGALRKALRGRAELVC LSGPHPVPDPPGPEGARSDFGSCPPEEQPRGWWFSEQEADVFSALEEPAVCRGLEESL GMVAQALNRLGPFDGLLGFSQGAALAALVCALGQAGDPRFPLPRFILLVSGFCPRGIG FKESILQRPLSLPSLHVFGDTDKVIPSQESVQLASQFPGAITLTHSGGHFIPAAAPQR QAYLKFLDQFAE
	SEQ ID NO: 75 987 bp
NOV15b, CG96412-03 DNA Sequence	ATGCCGGAAGGCCTCCTCCTCTTTGCCTGTACCATTGTGGATATCTTGGAAAGGTGTGT TGGATGACTTCACAGCGAGGGCCCTGGGAGCTGACTTCTTGGTGCACTACGGCCACAG TTGCCTGGTTCCCATGGACACCTCGGCCCAAGACTTCCGGGTGCTGTACGTCTTTGTG GACATCCGGATAGACACTACACACCTCCTGGACTCTCTCCGCCTCACCTTTCCCCCAG CCACTGCCCTTGCCCTGGTCAGCACCATTCAGTTTGTGTCGACCTTGCAGGCAGCCGC CCAGGAGCTGAAAGCCGAGTATCGTGTGAGTGTCCCACAGTGCAAGCCCCTGTCCCCT GGAGAGATCCTGGGCTGCACATCCCCCCGACTGTCCAAAGAGGTGGAGGCCGTTGTGG

	ACCCATATAGCAAAGTCCTATCCAGAGAACACTATGACCACCAGCGCATGCAGGCTGC TCGCCAAGAAGCCATAGCCACTGCCCGCTCAGCTAAGTCCTGGGGCCTTATTCTGGGC ACTTTGGGCCGCCAGGGCAGTCCTAAGATCCTGGAGCACCTGGAATCTCGACTCCGAG CCTTGGGCCTTTCCTTTGTGAGGCTGCTGCTCTCTGAGATCTTCCCCAGCAAGCTTAG CCTACTTCCTGAGGTGGATGTGGTGGCATGTCCACGTCTCTCCATTGACTGGGGCACA GCCTTCCCCAAGCCGCTGCTGACACCCTATGAGGCGGCCGTGGCTCTGAGGGACATTT CCTGGCAGCAGCCCTACCCGATGGACTTCTACGCTGGCAGCTCCTTGGGGCCCTGGAC GGTGAACCACGGCCAGGACCGCCGTCCCCACGCCCCGGGCCGGCCCGCGCGGGGGAAG GTACAGGAGGGGTCCGCGCGTCCCCCTTCGGCCGTGGCTTGCGAGGACTGCAGCTGCA GGGACGAGAAGGTGGCGCCGCTGGCTCCTTGACGCGCTCCCGGGCCTCAGGTATCAGC C
	ORF Start: ATG at 1 **ORF Stop: TGA at 958**
	SEQ ID NO: 76 **319 aa** **MW at 35041.8kD**
NOV15b, CG96412-03 Protein Sequence	MPEGLLLLFACTIVDILERCVDDFTARALGADFLVHYGHSCLVPMDTSAQDFRVLYVFV DIRIDTTHLLDSLRLTFPPATALALVSTIQFVSTLQAAAQELKAEYRVSVPQCKPLSP GEILGCTSPRLSKEVEAVVDPYSKVLSREHYDHQRMQAARQEAIATARSAKSWGLILG TLGRQGSPKILEHLESRLRALGLSFVRLLLSEIFPSKLSLLPEVDVVACPRLSIDWGT AFPKPLLTPYEAAVALRDISWQQPYPMDFYAGSSLGPWTVNHGQDRRPHAPGRPARGK VQEGSARPPSAVACEDCSCRDEKVAPLAP
	SEQ ID NO: 77 **977 bp**
NOV15c, 228116438 DNA Sequence	GCTTTTTATAAATGCCAACTTTGTACAAAAAAGCAGGCTCCGCGGCCGCCCCCTTCAC CATGCCGGAAGGCCTCCTCCTCTTTGCCTGTACCATTGTGGATATCTTGGAAAGGTTC ACGGAGGCCGAAGTGATGGTGATGGGTGACGTGACCTACGGGGCTTGCTGTGTGGATG ACTTCACAGCGAGGGCCCTGGGAGCTGACTTCTTGGTGCACTACGGCCACAGTTGCCT GATTCCCATGGACACCTCGGCCCAAGACTTCCGGGTGCTGTACGTCTTTGTGGACATC CGGATAGACACTACACACCTCCTGGACTCTCTCCGCCTCACCTTTCCCCCAGCCACTG CCCTTGCCCTGGTCAGCACCATTCAGTTTGTGTCGACCTTGCAGGCAGCCGCCCAGGA GCTGAAAGCCGAGTATCGTGTGAGTGTCCCACAGTGCAAGCCCCTGTCCCCTGGAGAG ATCCTGGGCTGCACATCCCCCGACTGTCCAGAGAGGTGGAGGCCGTTGTGTATCTTG GAGATGGCCGCTTCCATCTGGAGTCTGTCATGATTGCCAACCCCAATGTCCCCGCTTA CCGGTATGACCCATATAGCAAAGTCCTATCCAGAGAACACTATGACCACCAGCGCATG CAGGCTGCTCGCCAAGAAGCCATAGCCACTGCCCGCTCAGCTAAGTCCTGGGGCCTTA TTCTGGGCACTTTGGGCCGCCAGGGCAGTCCTAAGATCCTGGAGCACCTGGAATCTCG ACTCCGAGCCTTGGGCCTTTCCTTTGTGAGGCTGCTGCTCTCTGAGATCTTCCCCAGC AAGCTTAGCCTACTTCCCGAGGTGGATGTGTGGGTGCAGGTGGCATGTCCACGTCTCT CCATTGACTGGGGCACAGCCTTCCCCAAGCCGCTGCTGACACCCTATGAGGTAACACC AAGCTCTGGGAGAGAGTGGGCTTTGGACGTGGTTAAGGGTGGGCGCGCC
	ORF Start: ATG at 12 **ORF Stop: end of sequence**
	SEQ ID NO: 78 **322 aa** **MW at 35503.6kD**
NOV15c, 228116438 Protein Sequence	MPTLYKKAGSAAAPFTMPEGLLLLFACTIVDILERFTEAEVMVMGDVTYGACCVDDFTA RALGADFLVHYGHSCLIPMDTSAQDFRVLYVFVDIRIDTTHLLDSLRLTFPPATALAL VSTIQFVSTLQAAAQELKAEYRVSVPQCKPLSPGEILGCTSPRLSREVEAVVYLGDGR FHLESVMIANPNVPAYRYDPYSKVLSREHYDHQRMQAARQEAIATARSAKSWGLILGT LGRQGSPKILEHLESRLRALGLSFVRLLLSEIFPSKLSLLPEVDVWVQVACPRLSIDW GTAFPKPLLTPYEVTPSSGREWALDVVKGGRA
	SEQ ID NO: 79 **943 bp**
NOV15d, 228116442 DNA Sequence	AGGCTCCGCGGCCGCCCCCTTCACCATGCCGGAAGGCCTCCTCCTCTTTGCCTGTACC ATTGTGGATATCTTGGAAAGGTTCACGGAGGCCGAAGTGATGGTGATGGGTGACGTGA CCTACGGGGCTTGCTGTGTGGATGACTTCACAGCGAGGGCCCTGGGAGCTGACTTCTT GGTGCACTACGGCCACAGTTGCCTGATTCCCATGGACACCTCGGCCCAAGACTTCCGG GTGCTGTACGTCTTTGTGGACATCCGGATAGACACTACACACCTCCTGGACTCTCTCC GCCTCACCTTTCCCCCAGCCACTGCCCTTGCCCTGGTCAGCACCATTCAGTTTGTGTC AACCTTGCAGGCAGCCGCCCAGGAGCTGAAAGCCGAGTATCGTGTGAGTGTCCCACAG TGCAAGCCCCTGTCCCCTGGAGAGATCCTGGGCTGCACATCCCCCGACTGTCCAAAG AGGTGGAGGCCGTTGTGTATCTTGGAGATGGCCGCTTCCATCTGGAGTCTGTCATGAT

	TGCCAACCCCAATGTCCCCGCTTACCGGTATGACCCATATAGCAAAGTCCTATCCAGA GAACACTATGACCACCAGCGCATGCAGGCTGCTCGCCAAGAAGCCATAGCCACTGCCC GCTCAGCTAAGTCCTGGGGCCTTATTCTGGGCACTTTGGGCCGCCAGGGCAGTCCTAA GATCCTGGAGCACCTGGAATCTCGACTCCGAGCCTTGGGCCTTTCCTTTGTGAGGCTG CTGCTCTCTGAGATCTTCCCCAGCAAGCTTAGCCTACTTCCCGAGGTGGATGTGTGGG TGCAGGTGGCATGTCCACGTCTCTCCATTGACTGGGGCACAGCCTTCCCCAAGCCGCT GCTGACACCCTATGAGGTAACACCAAGCTCTGGGAGAGAGTGGGCTTTGGACGTGGTT AAGGGTGGGCGCGCC
	ORF Start: at 2	ORF Stop: end of sequence
	SEQ ID NO: 80	314 aa	MW at 34542.4kD
NOV15d, 228116442 Protein Sequence	GSAAAPFTMPEGLLLFACTIVDILERFTEAEVMVMGDVTYGACCVDDFTARALGADFL VHYGHSCLIPMDTSAQDFRVLYVFVDIRIDTTHLLDSLRLTFPPATALALVSTIQFVS TLQAAAQELKAEYRVSVPQCKPLSPGEILGCTSPRLSKEVEAVVYLGDGRFHLESVMI ANPNVPAYRYDPYSKVLSREHYDHQRMQAARQEAIATARSAKSWGLILGTLGRQGSPK ILEHLESRLRALGLSFVRLLLSEIFPSKLSLLPEVDVWVQVACPRLSIDWGTAFPKPL LTPYEVTPSSGREWALDVVKGGRA

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 15B.

Table 15B. Comparison of NOV15a against NOV15b through NOV15d.		
Protein Sequence	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV15b	1..188 1..135	134/188 (71%) 134/188 (71%)
NOV15c	1..390 17..321	302/390 (77%) 304/390 (77%)
NOV15d	1..390 9..313	303/390 (77%) 304/390 (77%)

Further analysis of the NOV15a protein yielded the following properties shown in Table 15C.

Table 15C. Protein Sequence Properties NOV15a	
PSort analysis:	0.7300 probability located in plasma membrane; 0.6400 probability located in endoplasmic reticulum (membrane); 0.2279 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 23 and 24

A search of the NOV15a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 15D.

Table 15D. Geneseq Results for NOV15a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAR99854	Human OVCA1 tumour suppressor protein - Homo sapiens, 538 aa. [WO9627609-A1, 12-SEP-1996]	1..371 81..367	284/371 (76%) 286/371 (76%)	e-149
ABB65850	Drosophila melanogaster polypeptide SEQ ID NO 24342 - Drosophila melanogaster, 454 aa. [WO200171042-A2, 27-SEP-2001]	1..380 83..380	184/383 (48%) 227/383 (59%)	6e-89
AAY43639	Amino acid sequecne of the DPH1 gene product - Saccharomyces cerevisiae, 425 aa. [WO9953762-A1, 28-OCT-1999]	1..373 99..390	157/377 (41%) 208/377 (54%)	7e-74
AAG41518	Arabidopsis thaliana protein fragment SEQ ID NO: 51665 - Arabidopsis thaliana, 453 aa. [EP1033405-A2, 06-SEP-2000]	1..380 70..373	159/386 (41%) 216/386 (55%)	6e-71
AAG41519	Arabidopsis thaliana protein fragment SEQ ID NO: 51666 - Arabidopsis thaliana, 378 aa. [EP1033405-A2, 06-SEP-2000]	7..380 1..298	154/380 (40%) 210/380 (54%)	1e-67

In a BLAST search of public sequence databases, the NOV15a protein was found to have homology to the proteins shown in the BLASTP data in Table 15E.

Table 15E. Public BLASTP Results for NOV15a				
Protein Accession Number	Protein/Organism/Length	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BTW7	DIPTHERIA TOXIN RESISTANCE PROTEIN REQUIRED FOR DIPHTHAMIDE BIOSYNTHESIS (SACCHAROMYCES)-LIKE 1 - Homo sapiens (Human), 363 aa.	1..371 1..287	285/371 (76%) 286/371 (76%)	e-149
Q9BZG8	CANDIDATE TUMOR SUPPRESSOR - Homo sapiens (Human), 443 aa.	1..371 81..367	285/371 (76%) 286/371 (76%)	e-149
Q16439	CANDIDATE TUMOR SUPPRESSOR PROTEIN - Homo sapiens (Human), 363 aa.	1..371 1..287	284/371 (76%) 285/371 (76%)	e-148
Q8WZ82	CANDIDATE TUMOR SUPPRESSOR OVCA2 - Homo sapiens (Human), 227 aa.	482..708 1..227	227/227 (100%) 227/227 (100%)	e-130
Q9D7E3	2310011M22RIK PROTEIN - Mus musculus (Mouse), 225 aa.	482..708 1..225	189/227 (83%) 205/227 (90%)	e-106

PFam analysis predicts that the NOV15a protein contains the domains shown in the Table 15F.

Table 15F. Domain Analysis of NOV15a			
Pfam Domain	NOV15a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Diphthamide_syn	1..385	172/421 (41%) 336/421 (80%)	3.6e-136
abhydrolase_2	504..707	42/249 (17%) 131/249 (53%)	0.47

5

EXAMPLE 16.

The NOV16 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 16A.

Table 16A. NOV16 Sequence Analysis		
	SEQ ID NO: 81	1492 bp
NOV16a, CG96511-01 DNA Sequence	ATGACCTGGGTACTCAGGACTCCTCCAGCACTCCTCCTGCTAGCTGTCATTGTTCTGG GCACATTTGGTAAGCCAACTGTGGCTTATGCTGAACTCCGCTGCATGTGTATAAAGAC AACCTCTGGAATTCATCCCAAAAACATCCAAAGTTTGGAAGTGATCGGGAAAGGAACC CATTGCAACCAAGTCGAAGTGATGGCCACACTGAAGGATGGGAGGAAAATCTGCCTGG ACCCAGATGCTCCCAGAATCAAGAAAATTGTAGGCTCTGTTTCCACAACCTTTCCCCA TTCTCTTGGCAGAGAACAGTCCCCACCTTCTTCAGCACTTATTTATATGCAGATACAA GAGCCAAAGTCATTCTTTCTCAGAGAAGGGAAGACTGTCCATATGGATCTCTTCCCCT TTCCTGCCAAAGTGGTCTCAGGGGATTTTTCTGACTTGGCCAATTCCAGCTCCATCAA TTTCCTGCGTGACTTCAGCCAAGTCACTCAACCTCTCTTAAACTCTATAATCTTATCC AGAAAAATGGGCAAAAATAGCTGTTGGGAGGATTAGGTAAATATAATTCATAAAACAC ACTTGGAACAAAGTAGTTGTCATGAAATATTAGTTGTTATTATATATGATAAGGTGCA GTGGGCTAGTGTGGAATTATGAGAGAAAAAGAAAAAAAAGAGATTTAGAAAACTCAAG ACATTGAGCTTATTCTGCCACAAATGTTTAGATAAGGAAGAGCTTTCTTGTCTGTTAA AATACTTATTTAGAGTAAAGGGCAAGTACAATACCAACAACTATAATCTTTCTAGTGT TGACCTAGCTAGACCTTTATATTTCATGAACAGAGGTAAGATTAATCAGATGAAATGC CAAATGAAAAGAGGAAGACAGTTGAGTAAAACAAAAAGTGTGATTTAGCAGAGGATTT AGTGCGTCAAGTTATACCACATCCTGTTAGTCTTTTATGCCGTTATCCAGTGTTGTGT GTCCCAGCAGCATAACAAAGGTTAAGAACAAGTTTATTCAACATCAGCAATGATTGCC AGAGGGAAGCAACTCACTTTAGTCAACATGGCTATCAAATGACAAAGAAAGGCACAAG TGAGAAATCTGTTTACAACCCTGGCATTTGGAGACACAACAGATGAATTATAGCATTT TGTAGATATTGAATTATATTTTTACATTCACTCACAATACTTGGAGTGAAGAGATCAA AAATTCTGTTAGAACAAATTGTAATGCATTTTCTCTCTTGTTTCTATTTTTCCCCACT TCTCCAATATACTTACATACTTTTAATAGGTCGGTAAATCTCCTAAATATTAGCTCAC ATCCAATCATACTGTTTAGAGAGCTTCTATGATTCATAAACATGCCTGAATGGATTTG GAGCGAATTCACAATTTCAGTATTTACCAAAATGTTAAATACTATGATTTGTATCACA GGTATCAACAGAGCTAAAGTAGCCAGATTAGTAAAACTTCTC
	ORF Start: ATG at 1	ORF Stop: TAG at 556
	SEQ ID NO: 82	185 aa	MW at 20434.7kD
NOV16a, CG96511-01 Protein Sequence	MTWVLRTPPALLLLAVIVLGTFGKPTVAYAELRCMCIKTTSGIHPKNIQSLEVIGKGT HCNQVEVMATLKDGRKICLDPDAPRIKKIVGSVSTTFPHSLGREQSPPSSALIYMQIQ EPKSFFLREGKTVHMDLFPFPAKVVSGDFSDLANSSSINFLRDFSQVTQPLLNSIILS RKMGKNSCWED	

Further analysis of the NOV16a protein yielded the following properties shown in

Table 16B.

Table 16B. Protein Sequence Properties NOV16a	
PSort analysis:	0.7427 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Cleavage site between residues 24 and 25

A search of the NOV16a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 16C.

	Table 16C. Geneseq Results for NOV16a			
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB07712	Amino acid sequence of platelet basic protein - Homo sapiens, 94 aa. [WO200042069-A1, 20-JUL-2000]	29..88 24..83	59/60 (98%) 60/60 (99%)	6e-29
AAB07711	Amino acid sequence of connective tissue-activating peptide - Homo sapiens, 85 aa. [WO200042069-A1, 20-JUL-2000]	29..88 15..74	59/60 (98%) 60/60 (99%)	6e-29
AAB07710	Amino acid sequence of beta-thromboglobulin - Homo sapiens, 81 aa. [WO200042069-A1, 20-JUL-2000]	29..88 11..70	59/60 (98%) 60/60 (99%)	6e-29
AAB07709	Amino acid sequence of neutrophil activating peptide 2 (NAP-2) variant - Homo sapiens, 73 aa. [WO200042069-A1, 20-JUL-2000]	29..88 3..62	59/60 (98%) 60/60 (99%)	6e-29
AAB07708	Amino acid sequence of neutrophil activating peptide 2 (NAP-2) variant - Homo sapiens, 74 aa. [WO200042069-A1, 20-JUL-2000]	29..88 4..63	59/60 (98%) 60/60 (99%)	6e-29

5

In a BLAST search of public sequence databases, the NOV16a protein was found to have homology to the proteins shown in the BLASTP data in Table 16D.

Table 16D. Public BLASTP Results for NOV16a				
Protein Accession Number	Protein/Organism/Length	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAA72500	CONNECTIVE TISSUE ACTIVATING PEPTIDE-III - synthetic construct, 91 aa (fragment).	29..88 21..80	59/60 (98%) 60/60 (99%)	2e-28
P02775	Platelet basic protein precursor (PBP) [Contains: Connective-tissue activating peptide III (CTAP-III); Low-affinity platelet factor IV (LA-PF4); Beta-thromboglobulin (Beta-TG); Neutrophil-activating peptide 2 (NAP-2)] - Homo sapiens (Human), 128 aa.	29..88 58..117	59/60 (98%) 60/60 (99%)	2e-28
AAA73218	COL-CTAP-III(LEU21) FUSION PROTEIN - synthetic construct, 591 aa.	29..88 521..580	58/60 (96%) 60/60 (99%)	3e-28
AAA73217	CONNECTIVE TISSUE ACTIVATING PEPTIDE III PRECURSOR - synthetic construct, 91 aa.	29..88 21..80	58/60 (96%) 60/60 (99%)	3e-28
AAA73216	CTAP-III(LEU21)--HIRUDIN FUSION PROTEIN PRECURSOR - synthetic construct, 162 aa.	29..88 21..80	57/60 (95%) 60/60 (100%)	1e-27

PFam analysis predicts that the NOV16a protein contains the domains shown in the Table 16E.

Table 16E. Domain Analysis of NOV16a			
Pfam Domain	NOV16a Match Region	Identities/ Similarities for the Matched Region	Expect Value
IL8	24..92	34/70 (49%) 63/70 (90%)	1.6e-32

5

EXAMPLE 17.

The NOV17 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 17A.

Table 17A. NOV17 Sequence Analysis	
SEQ ID NO: 83	3030 bp
NOV17a, CG96522-01 DNA Sequence	CCTGGCTCTCCCCTATGGTTCTTCTTTCTTTAGCTTTGGCCCCGGGGGGCCACGGCAGA CCTGGCAGGGCGGGCGGCACAGGGGAGAGGCAGGAAGAAACTGGATCCCTGGGGACCT GTGGTAGGGTCGGCAGGAAGAAACTGGATCCCTGGGGACCTGTGGCCACGGGCCCTCC CTGAGCACCGCGCGCAAAGGCCCGGCCCCAGGGCCAGGCAACTCCAGCGCCGAGGCCG TCCAGTGCGGGGCCAGGCCCCGGGGGTGCCCTGCTGCACCGACTCCGGGCGGAGATTG GCTGTCCTCGGCCCGCAATGCTGTGCGCCCCGCCCCGCGCGCCGCTTCCGGCGGAGTC AGGCTTGGCTAATCGAGCGCGCGGGCTGGCGGCTCGGCGTTCGTTTGGCCGCGCCTGC CCGTGTGGTGGTTTCCGGCGGAGGTGGTGAGAGCCGGCGGGCAGGTGGGCTTGGCCGC GCTGTGGGTGCCTGGGACCCGCAGGGAGGATGGGCGCGGTGGCGCGGCCTGGCGGGGG GCTCGTCTCCGGGGTCCCCGGGTCCTGGTGAGAGCGGGGTCCCTCGACGCCGTGGCGG TCTCGAACCTGTGGATCTGAGGAGGGGATGCACACACAGCAGCCAGCCCAGTGTGGTG CCGAGAAACAGAGCCCCGAGGCCCTGGTCCTCAGAAAGGTCCCTCCCCTGCCTTCCTG TCCCTGCAGAGGTCATGCAGAAATTCTCTGGCTGGCCCGAAGTCCAGCTCAGGGCCAT GAAGAGGCTTGTGGCCGTGGGCCCCGATGTCTTCCAGGCTCACCAGGAGGACACAGAG CGCTATGTGCTCACCAACCTCAACATCGGGGCAGAACTGCTTCGGGACCCGTCCCTGG GGGCTCAGTTTCGGGTGCACCTGGTGAAGATGGTCATTCTGACAGAGCCTGAGGGTGC CCCAAATATCACAGCCAACCTCACCTCGTCCCTGCTGAGCGTCTGTGGGTGGAGCCAG ACCATCAACCCTGAGGACGACACGGATCCTGGCCATGCTGACCTGGTCCTCTATATCA CTAGGAGGTTTGACCTGGAGTTGCCTGATGGTAACCGGCAGGTGCGGGGCGTCACCCA GCTGGGCGGTGCCTGCTCCCCAACCTGGAGCTGCCTCATTACCGAGGACACTGGCTTC GACCTGGGAGTCACCATTGCCCATGAGATTGGGCACAGGTATGTAGCCCCACCAGCTG TCCCCAGGATCTGGCAAGGAGCTGACCTGGGTACCCAGGGTGGAGGTGGTCTTAGCAA GCAGTGGGTCCTTGTAGAGTTTCTCCAGAGGAGCCTGTACCCCTCACCCCGACAGACT CAGGTGAGCTTCGGCCTGGAGCACGACGGCGCGCCCGGCAGCGGCTGCGGCCCCAGCG GACACGTGATGGCTTCGGACGGCGCCGCGCCCCGCGCCGGCCTCGCCTGGTCCCCCTG CAGCCGCCGGCAGCTGCTGAGCCTGCTCGGACGGGCGCGCTGCGTGTGGGACCCGCCG CGGCCTCAACCCGGGTCCGCGGGGCACCCGCCGGATGCGCAGCCTGGCCTCTACTACA GCGCCAACGAGCAGTGCCGCGTGGCCTTCGGCCCCAAGGCTGTCGCCTGCGATATGTG CCAGGCCCTCTCCTGCCACACAGACCCGCTGGACCAAAGCAGCTGCAGCCGCCTCCTC GTTCCTCTCCTGGATGGGACAGAATGTGGCGTGGAGAAGTGGTGCTCCAAGGGTCGCT GCCGCTCCCTGGTGGAGCTGACCCCCATAGCAGCAGTGCATGGGCGCTGGTCTAGCTG GGGTCCCCGAAGTCCTTGCTCCCGCTCCTGCGGAGGAGGTGTGGTCACCAGGAGGCGG CAGTGCAACAACCCCAGGTACCGCAGGGAGGGTGCTTTTCTGTCAGGGTGTCCTGGGG GGAAGCCGGAAGTGAGTCACAGTCAGCTCTTCCGAGCCTCCAGTGTGCACGCCTGTAA GCTGGGATCGGTCCTCAGCGATGTCCATCAGTGCAGACACATGTGCCGGGCCATTGGC GAGAGCTTCATCATGAAGCGTGGAGACAGCTTCCTCGATGGGACCCGGTGTATGCCAA GTGGCCCCCGGGAGGACGGGACCCTGAGCCTGTGTGTGTCGGGCAGCTGCAGGGTAGG CGGCTGTGATGGTAGGATGGACTCCCAGCAGGTATGGGACAGGTGCCAGGTGTGTGGT GGGGACAACAGCACGTGCCACGGCGTCGAGGGACCCCGCTCTCACCAGGACCCGGGGA CCCCGGAGACGAGCCCCCGCCAGGCCGCGCCACCGCGCCCATCCTCCCTGCGGGTCC CAGGCAGGCTTGCGGCACTGGCCAGATGTGGGCATCGAGGGGGCAGGTGCGGAATGTC ACCACCTCTCCCATACCCGCAAGGCCGATCTGCCTTCAGCTCCCAGCAAGTGTGGGGC AGCGCGGGCCACAGAGTAGGGTGCAGGGATGGGGCCCCGGGAGGAGCAGGCCCACCTC ACTGAACTCTATCCCAGACAGCCTGCCTAGCACAACACAGGGAGGCCCCCAAATGGCT CATTCCTCAGCCATCAGCAGCAGCCTGCATAGAGGACACTGGGGTTACCAGGGGATGG TTACCTGGTCCCCAAATCACCTTGTTGTGGCAAGTGCCAGAATTCCTAAGCCACGGCA GGCCTGGGTGTGGGCCGCTGTGCGTGGGCCCTGCTCGGTGAGCTGTGGGGCAGGTGAG ACCTGGGGAAGGCTCATCCACAGCACGGCTTGCGTGGAGGCCCAGGGCAGCCTCCTGA AGACATTGCCCCCAGCCCGGTGCAGAGCAGGGGCCCAGCAGCCAGCTGTGGCGCTGGA AACCTGCAACTCTGAGTCTCAATTTCCCATCTGTGAAATGGAGATAATAGCAGTAGGT

	CCCTCCCTGGGCGCTACAAGGATTCAGGGAGATAATCGGAAAATGCCAAGTGTGTTCC TTGGTTCATGATACTTTTTTTTGTGAGACAGAGTCTTGCTCTGTCGCCCAGGCTGGAGT GCAGGGGCGTAATC

	ORF Start: ATG at 15		ORF Stop: TGA at 2967
	SEQ ID NO: 84	984 aa	MW at 104943.2kD
NOV17a, CG96522-01 Protein Sequence	MVLLSLALAPGATADLAGRAAQGRGRKKLDPWGPVVGSAGRNWIPGDLWPRALPEHRA QRPGPRARQLQRRGRPVRGQAPGVPCCTDSGRRLAVLGPQCCAPRPARRFRRSQAWLI ERAGWRLGVRLAAPARVVVSGGGGESRRAGGLGRAVGAWDPQGGWARWRGLAGGSSPG SPGPGESGVPRRRGGLEPVDLRRGCTHSSQPSVVPRNRAPRPWSSERSLPCLPVPAEV MQKFSGWPEVQLRAMKRLVAVGPDVFQAHQEDTERYVLTNLNIGAELLRDPSLGAQFR VHLVKMVILTEPEGAPNITANLTSSLLSVCGWSQTINPEDDTDPGHADLVLYITRRFD LELPDGNRQVRGVTQLGGACSPTWSCLITEDTGFDLGVTIAHEIGHRYVAPPAVPRIW QGADLGTQGGGGLSKQWVLVEFLQRSLYPSPRQTQVSFGLEHDGAPGSGCGPSGHVMA SDGAAPRAGLAWSPCSRRQLLSLLGRARCVWDPPRPQPGSAGHPPDAQPGLYYSANEQ CRVAFGPKAVACDMCQALSCHTDPLDQSSCSRLLVPLLDGTECGVEKWCSKGRCRSLV ELTPIAAVHGRWSSWGPRSPCSRSCGGGVVTRRRQCNNPRYRREGAFLSGCPGGKPEV SHSQLFRASSVHACKLGSVLSDVHQCRHMCRAIGESFIMKRGDSFLDGTRCMPSGPRE DGTLSLCVSGSCRVGGCDGRMDSQQVWDRCQVCGGDNSTCHGVEGPRSHQDPGTPETS PPPGRATAPILPAGPRQACGTGQMWASRGQVRNVTTSPIPARPICLQLPASVGQRGPQ SRVQGWGPGRSRPTSLNSIPDSLPSTTQGGPQMAHSSAISSSLHRGHWGYQGMVTWSP NHLVVASARIPKPRQAWVWAAVRGPCSVSCGAGETWGRLIHSTACVEAQGSLLKTLPP ARCRAGAQQPAVALETCNSESQFPICEMEIIAVGPSLGATRIQGDNRKMPSVFLGS

Further analysis of the NOV17a protein yielded the following properties shown in Table 17B.

Table 17B. Protein Sequence Properties NOV17a	
PSort analysis:	0.5500 probability located in lysosome (lumen); 0.3700 probability located in outside; 0.1132 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 15 and 16

A search of the NOV17a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 17C.

Table 17C. Geneseq Results for NOV17a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB04153	Human ADAMTS-M polypeptide - Homo sapiens, 1416 aa. [EP1152055-A1, 07-NOV-2001]	192..948 53..825	492/860 (57%) 526/860 (60%)	0.0
AAG63829	Amino acid sequence of a human zdint5 polypeptide - Homo sapiens, 1120 aa. [WO200159112-A1, 16-AUG-2001]	192..741 60..541	341/559 (61%) 368/559 (65%)	e-175
AAG63826	Amino acid sequence of a human zdint5 polypeptide - Homo sapiens, 203 aa. [WO200159112-A1, 16-AUG-2001]	247..486 9..199	187/240 (77%) 190/240 (78%)	e-100
AAU72894	Human metalloprotease partial protein sequence #6 - Homo sapiens, 1428 aa. [WO200183782-A2, 08-NOV-2001]	247..771 641..1133	178/563 (31%) 237/563 (41%)	1e-59
AAB42668	Human ORFX ORF2432 polypeptide sequence SEQ ID NO:4864 - Homo sapiens, 118 aa. [WO200058473-A2, 05-OCT-2000]	285..396 1..111	106/112 (94%) 107/112 (94%)	4e-56

5

In a BLAST search of public sequence databases, the NOV17a protein was found to have homology to the proteins shown in the BLASTP data in Table 17D.

Table 17D. Public BLASTP Results for NOV17a				
Protein Accession Number	Protein/Organism/Length	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAD12729	ADAMTS-13 PROTEIN, VARIANT 2 - Homo sapiens (Human), 1371 aa.	192..948 30..802	492/860 (57%) 526/860 (60%)	0.0
BAB69487	VON WILLEBRAND FACTOR-CLEAVING PROTEASE - Homo sapiens (Human), 1427 aa.	192..948 30..802	492/860 (57%) 526/860 (60%)	0.0
Q96L37	VON WILLEBRAND FACTOR-CLEAVING PROTEASE PRECURSOR - Homo sapiens (Human), 1427 aa.	192..948 30..802	492/860 (57%) 526/860 (60%)	0.0
CAC83682	ADAMTS-13 PROTEIN - Homo sapiens (Human), 1340 aa.	192..948 30..771	463/858 (53%) 497/858 (56%)	0.0
CAC69385	SEQUENCE 10 FROM PATENT WO0159112 - Homo sapiens (Human), 1118 aa (fragment).	192..741 60..541	341/559 (61%) 368/559 (65%)	e-174

PFam analysis predicts that the NOV17a protein contains the domains shown in the Table 17E.

Table 17E. Domain Analysis of NOV17a			
Pfam Domain	NOV17a Match Region	Identities/ Similarities for the Matched Region	Expect Value
tsp_1	593..652	20/61 (33%) 44/61 (72%)	0.011

5

EXAMPLE 18.

The NOV18 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 18A.

Table 18A. NOV18 Sequence Analysis		
	SEQ ID NO: 85	1103 bp
NOV18a, CG96535-01 DNA Sequence	ATGCATTAGGAAGATCCTGGACCTAGAGAACAAGTCCCCCGAACGCTGAGTTGGAGGC GGGACTTCGGGTGCGCGTTGGCGGGAGCATGCTGGGGCTCTGGGGGCAGCGGCTCCCC GCGGCGTGGGTCCTGCTTCTGTTGCCTTTCCTGCCGCTGCTGCTGCTTGCAGCCCCCG CGCCCCACCGCGCGCCCTACAAGCCGGTCATCGTGGTGCATGGGCTCTTCGACAGCTC GTACAGCTTCCGCCACCTGCTGGAATACATCAATGAGACACACCCCGGGACTGTGGTG ACAGTGCTCGATCTCTTCGATGGGAGAGAGAGCTTGCGACCCCTGTGGGAACAGGTGC AAGGGTTCCGAGAGGCTGTGGTCCCCATCATGGCAAAGGCCCCTCAAGGGGTGCATCT CATCTGCTACTCGCAGGGGGGCCTTGTGTGCCGGGCTCTGCTTTCTGTCATGGATGAT CACAACGTGGATTCTTTCATCTCCCTCTCCTCTCCACAGATGGGACAGTATGGAGACA CGGACTACTTGAAGTGGCTGTTCCCCACCTCCATGCAGTCTAACCTCTATCGGATCTG CTATAGCCCCTGGGGCCAGGAATTCTCCATCTGCAACTACTGGCATGATCCCCACCAC GATGACTTGTACCTCAATGCCAGCAGCTTCCTGGCCCTGATCAATGGGGAAAGAGACC ATCCCAATGCCGCAGTATGGCGGAAGAACTTTCTGCGTGTGGGCCACCTGGTGCTGAT TGGGGGCCCTGATGATGGTGTTATTACTCCCTGGCAGTCCAGCTTCTTTGGTTTCTAT GATGCAAATGAGACCGTCCTGGAGATGGAGGAGCAACTGCCTGCCAGGCCCACCCACC AGTCTGAGCTGCTTCTGCTGAGGCTGGTCTGCTTGAAGCCTCCCAGGAGAAAGAAGCC AGGTGGGAATGGAGAGAGAGAGGAAGCCTGTAGGGTCCAGCGTCAAAGCGAATCATGG GGCCCAGGGCTGAGCTGTGCACTCTCTTAGGCGGATTCTCCTTCCTCCTGCTACTGAT ACCAGGCGAGGGGGCCAAGGGTGGATCCCTCAGAGAGAGGTGACAACAGAGGGGGTAG G
	ORF Start: ATG at 87	ORF Stop: TAG at 1014
	SEQ ID NO: 86	309 aa	MW at 34918.7kD
NOV18a, CG96535-01 Protein Sequence	MLGLWGQRLPAAWVLLLLPFLPLLLLAAPAPHRAPYKPVIVVHGLFDSSYSFRHLLEY INETHPGTVVTVLDLFDGRESLRPLWEQVQGFREAVVPIMAKAPQGVHLICYSQGGLV CRALLSVMDDHNVDSFISLSSPQMGQYGDTDYLKWLFPTSMQSNLYRICYSPWGQEFS ICNYWHDPHHDDLYLNASSFLALINGERDHPNAAVWRKNFLRVGHLVLIGGPDDGVIT PWQSSFFGFYDANETVLEMEEQLPARPTHQSELLLLRLVCLKPPRRKKPGGNGEREEA CRVQRQSESWGPGLSCALS	
	SEQ ID NO: 87	1103 bp
NOV18b, CG96535-02 DNA Sequence	ATGCATTAGGAAGATCCTGGACCTAGAGAACAAGTCCCCCGAACGCTGAGTTGGAGGC GGGACTTCGGGTGCGCGTTGGCGGGAGCATGCTGGGGCTCTGGGGGCAGCGGCTCCCC GCGGCGTGGGTCCTGCTTCTGTTGCCTTTCCTGCCGCTGCTGCTGCTTGCAGCCCCCG CGCCCCACCGCGCGTCCTACAAGCCGGTCATCGTGGTGCATGGGCTCTTCGACAGCTC GTACAGCTTCCGCCACCTGCTGGAATACATCAATGAGACACACCCCGGGACTGTGGTG ACAGTGCTCGATCTCTTCGATGGGAGAGAGAGCTTGCGACCCCTGTGGGAACAGGTGC AAGGGTTCCGAGAGGCTGTGGTCCCCATCATGGTAAAGGCCCCTCAAGGGGTGCATCT CATCTGCTACTCGCAGGGGGGCCTTGTGTGCCGGGCTCTGCTTTCTGTCATGGATGAT CACAACGTGGATTCTTTCATCTCCCTCTCCTCTCCACAGATGGGACAGTATGGAGACA CGGACTACTTGAAGTGGCTGTTCCCCACCTCCATGCGGTCTAACCTCTATCGGATCTG CTATAGCCCCTGGGGCCAGGAATTCTCCATCTGCAACTACTGGCATGATCCCCACCAC GATGACTTGTACCTCAATGCCAGCAGCTTCCTGGCCCTGATCAATGGGGAAAGAGACC ATCCCAATGCCGCAGTATGGCGGAAGAACTTTCTGCGTGTGGGCCACCTGGTGCTGAT TGGGGGCCCTGATGATGGTGTTATTACTCCCTGGCAGTCCAGCTTCTTTGGTTTCTAT GATGCAAATGAGACCGTCCTGGAGATGGAGGAGCAACTGCCTGCCAGGCCCACCCACC AGTCTGAGCTGCTTCTGCTGAGGCTGGTCTGCTTGAAGCCTCCCAGGAGAAAGAAGCC AGGTGGGAATGGAGAGAGAGAGGAAGCCTGTAGGGTCCAGCGTCAAAGCGAATCATGG GGCCCAGGGCTGAGCTGTGCACTCTCTTAGGCGGATTCTCCTTCCTCCTGCTACTGAT

	ACCAGGCGAGGGGGCCAAGGGTGGATCCCTCAGAGAGAGGTGACAACAGAGGGGGTAG G		
	ORF Start: ATG at 87	ORF Stop: TAG at 1014	
	SEQ ID NO: 88	309 aa	MW at 34964.7kD
NOV18b, CG96535-02 Protein Sequence	MLGLWGQRLPAAWVLLLLPFLPLLLLAAPAPHRASYKPVIVVHGLFDSSYSFRHLLEY INETHPGTVVTVLDLFDGRESLRPLWEQVQGFREAVVPIMVKAPQGVHLICYSQGGLV CRALLSVMDDHNVDSFISLSSPQMGQYGDTDYLKWLFPTSMRSNLYRICYSPWGQEFS ICNYWHDPHHDDLYLNASSFLALINGERDHPNAAVWRKNFLRVGHLVLIGGPDDGVIT PWQSSFFGFYDANETVLEMEEQLPARPTHQSELLLLRLVCLKPPRRKKPGGNGEREEA CRVQRQSESWGPGLSCALS		

Sequence comparison of the above protein sequences yields the following

sequence relationships shown in Table 18B.

Table 18B. Comparison of NOV18a against NOV18b.		
Protein Sequence	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV18b	1..309 1..309	266/309 (86%) 267/309 (86%)

Further analysis of the NOV18a protein yielded the following properties shown in

Table 18C.

Table 18C. Protein Sequence Properties NOV18a	
PSort analysis:	0.8200 probability located in outside; 0.6850 probability located in plasma membrane; 0.4882 probability located in lysosome (lumen); 0.1370 probability located in microbody (peroxisome)
SignalP analysis:	Cleavage site between residues 28 and 29

5

A search of the NOV18a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 18D.

Table 18D. Geneseq Results for NOV18a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV18a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAG89194	Human secreted protein, SEQ ID NO: 314 - Homo sapiens, 280 aa. [WO200142451-A2, 14-JUN-2001]	1..255 7..233	224/255 (87%) 225/255 (87%)	e-128
AAG89195	Human secreted protein, SEQ ID NO: 315 - Homo sapiens, 174 aa. [WO200142451-A2, 14-JUN-2001]	1..168 7..174	166/168 (98%) 167/168 (98%)	2e-95
ABB61020	Drosophila melanogaster polypeptide SEQ ID NO 9852 - Drosophila melanogaster, 288 aa. [WO200171042-A2, 27-SEP-2001]	16..249 3..235	100/234 (42%) 138/234 (58%)	2e-49
AAB56711	Human prostate cancer antigen protein sequence SEQ ID NO:1289 - Homo sapiens, 318 aa. [WO200055174-A1, 21-SEP-2000]	5..255 8..275	73/276 (26%) 121/276 (43%)	4e-12
AAG45320	Arabidopsis thaliana protein fragment SEQ ID NO: 56883 - Arabidopsis thaliana, 304 aa. [EP1033405-A2, 06-SEP-2000]	38..254 27..243	63/226 (27%) 99/226 (42%)	3e-09

5

In a BLAST search of public sequence databases, the NOV18a protein was found to have homology to the proteins shown in the BLASTP data in Table 18E.

Table 18E. Public BLASTP Results for NOV18a				
Protein Accession Number	Protein/Organism/Length	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
T14739	hypothetical protein DKFZp564P1516.1 - human, 308 aa.	1..255 7..261	252/255 (98%) 253/255 (98%)	e-152
Q9UMR5	Palmitoyl-protein thioesterase 2 precursor (EC 3.1.2.22) (Palmitoyl-protein hydrolase 2) (PPT-2) (G14) - Homo sapiens (Human), 302 aa.	1..255 1..255	252/255 (98%) 253/255 (98%)	e-152
O70489	Palmitoyl-protein thioesterase 2 precursor (EC 3.1.2.22) (Palmitoyl-protein hydrolase 2) (PPT-2) - Rattus norvegicus (Rat), 302 aa.	1..254 1..254	239/254 (94%) 244/254 (95%)	e-144
O35448	Palmitoyl-protein thioesterase 2 precursor (EC 3.1.2.22) (Palmitoyl-protein hydrolase 2) (PPT-2) - Mus musculus (Mouse), 302 aa.	1..254 1..254	235/254 (92%) 242/254 (94%)	e-142
Q9VKH6	CG4851 PROTEIN - Drosophila melanogaster (Fruit fly), 288 aa.	16..249 3..235	100/234 (42%) 138/234 (58%)	4e-49

PFam analysis predicts that the NOV18a protein contains the domains shown in the Table 18F.

Table 18F. Domain Analysis of NOV18a			
Pfam Domain	NOV18a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Palm_thioest	177..242	26/66 (39%) 41/66 (62%)	4.1e-07

5

179

EXAMPLE 19.

The NOV19 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 19A.

Table 19A. NOV19 Sequence Analysis		
	SEQ ID NO: 89	525 bp
NOV19a, CG96567-02 DNA Sequence	GAAGGAGGGAGACTTGTCTAGGGGCTGCCCGGCCCGGCAGAGCGGGGTTGATGGACCG GGCCGCCCGGTGCAGCGGCGCCAGCTCCCTGCCACTGCTCCTGGCCCTTGCCCTGGGT CTAGTGATCCTTCACTGTGTGGTGGCAGATGGGAATTCCACCAGAAGTCCTGAAACTA ATGGCCTCCTCTGTGGAGACCCTGAGGAAAACTGTGCAGCTACCACCACACAATCAAA GCGGAAAGGCCACTTCTCTAGGTGCCCCAAGCAATACAAGCATTACTGCATCAAAGGG AGATGCCGCTTCGTGGTGGCCGAGCAGACGCCCTCCTGTGTCCCTCTTCGGAAACGTC GTAAAGAAAGAAGAAAGAAGAAGAAATGGAAACTCTGGGTAAAGATATAACTCCTAT CAATGAAGATATTGAAGAGACAAATATTGCT**TAA**AAGGCTATGAAGTTACCTCCAGGT TGGTGGCAAGCTGCAAAGTGCCTTGCTCATTTGAAAATGGACAGAATGTGTCTCAGGA AAA	
	ORF Start: ATG at 51	ORF Stop: TAA at 438
	SEQ ID NO: 90	129 aa	MW at 14301.3kD
NOV19a, CG96567-02 Protein Sequence	MDRAARCSGASSLPLLLALALGLVILHCVVADGNSTRSPETNGLLCGDPEENCAATTT QSKRKGHFSRCPKQYKHYCIKGRCRFVVAEQTPSCVPLRKRRKRKKKEEEMETLGKDI TPINEDIEETNIA	

Further analysis of the NOV19a protein yielded the following properties shown in Table 19B.

Table 19B. Protein Sequence Properties NOV19a	
PSort analysis:	0.8200 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Cleavage site between residues 32 and 33

A search of the NOV19a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 19C.

Table 19C. Geneseq Results for NOV19a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU03570	Human betacellulin growth factor splice variant BTC-beta polypeptide - Homo sapiens, 129 aa. [WO200149845-A1, 12-JUL-2001]	1..129 1..129	129/129 (100%) 129/129 (100%)	3e-72
AAR40168	Growth factor BTC-GF of human - Homo sapiens, 178 aa. [EP555785-A, 18-AUG-1993]	1..129 1..178	129/178 (72%) 129/178 (72%)	4e-65
AAR40167	Recombinant growth factor BTC-GF of mouse - Mus musculus, 177 aa. [EP555785-A, 18-AUG-1993]	1..129 1..177	99/178 (55%) 106/178 (58%)	9e-45
AAY50768	Non-human animal beta-caerulein protein 1 - Unidentified, 80 aa. [JP11285332-A, 19-OCT-1999]	32..94 1..63	63/63 (100%) 63/63 (100%)	8e-34
AAB03140	Human betacellulin (BTC) protein - Homo sapiens, 80 aa. [WO200025803-A1, 11-MAY-2000]	32..94 1..63	63/63 (100%) 63/63 (100%)	8e-34

5

In a BLAST search of public sequence databases, the NOV19a protein was found to have homology to the proteins shown in the BLASTP data in Table 19D.

Table 19D. Public BLASTP Results for NOV19a				
Protein Accession Number	Protein/Organism/Length	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96F48	SIMILAR TO BETACELLULIN - Homo sapiens (Human), 178 aa.	1..129 1..178	129/178 (72%) 129/178 (72%)	9e-65
P35070	Betacellulin precursor (BTC) - Homo sapiens (Human), 178 aa.	1..129 1..178	129/178 (72%) 129/178 (72%)	9e-65
Q9TTC5	Betacellulin precursor (BTC) - Bos taurus (Bovine), 178 aa.	1..129 1..178	113/178 (63%) 119/178 (66%)	6e-54
Q9JJM4	BETACELLULIN PRECURSOR - Rattus norvegicus (Rat), 177 aa.	1..129 1..177	99/178 (55%) 107/178 (59%)	1e-44
Q05928	Betacellulin precursor (BTC) - Mus musculus (Mouse), 177 aa.	1..129 1..177	99/178 (55%) 106/178 (58%)	2e-44

PFam analysis predicts that the NOV19a protein contains the domains shown in the Table 19E.

Table 19E. Domain Analysis of NOV19a			
Pfam Domain	NOV19a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

182

EXAMPLE 20.

The NOV20 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 20A.

Table 20A. NOV20 Sequence Analysis		
	SEQ ID NO: 91	465 bp
NOV20a, CG96637-01 DNA Sequence	CCTGCCTGCCCACCAGGAGG**ATGA**AGGTCTCCGTGGCTGCCCTCTCCTGCCTCATGCT TGTTACTGCCCTTGGATCCCAGGCCCGGGTCACAAAAGGTGAGTCCAGTGAGCTTGGT CCTCAGATGACCCTTTCTCATGCTGCAGGATTCCATGCTACTAGTGCTGACTGCTGCA TCTCCTACACCCCACGAAGCATCCCGTGTTCACTCCTGGAGAGTTACTTTGAAACGAA CAGCGAGTGCTCCAAGCCGGGTGTCATGTTCCTCACCAAGAAGGGGCGACGTTTCTGT GCCAACCCCAGTGATAAGCAAGTTCAGGTTTGCGTGAGAATGCTGAAGCTGGACACAC GGATCAAGACCAGGAAGAAT**TGA**ACTTGTCAAGGTGAAGGGACACAAGTTGCCAGCCA CCAACTTTCTTGCCTCAACTACCTTCCTGAATTATTTTTTAAAGAAGCATTTATTCTT G	
	ORF Start: ATG at 21	ORF Stop: TGA at 369
	SEQ ID NO: 92	116 aa MW at 12651.6kD
NOV20a, CG96637-01 Protein Sequence	MKVSVAALSCLMLVTALGSQARVTKGESSELGPQMTLSHAAGFHATSADCCISYTPRS IPCSLLESYFETNSECSKPGVMFLTKKGRRFCANPSDKQVQVCVRMLKLDTRIKTRKN	
	SEQ ID NO: 93	387 bp
NOV20b, CG96637-04 DNA Sequence	GCAGTGAGCCCAGGAGTCCTCGGCCAGCCCTGCCTGCCCACCAGGAGG**ATGA**AGGTCT CCGTGGCTGCCCTCTCCTGCCTCATGCTTGTTACTGCCCTTGGATCCCAGGCCCGGGT CACAAAAGATACAGAGACAGAGTTCATGATGTCAAAGCTTCCATTGGAAAATCCAGTA CTTCTGGACAGATTCCATGCTACTAGTGCTGACTGCTGCATCTTCCTCACCAAGAAGG GGCGACGTTTCTGTGCCAACCCCAGTGATAAGCAAGTTCAGGTTTGCATGAGAATGCT GAAGCTGGACACACGGATCAAGACCAGGAAGAAT**TGA**ACTTGTCAAGGTGAAGGGACA CAAGTTGCCAGCCACCAACTTTCTTGCCTCAACTAAAGG	
	ORF Start: ATG at 49	ORF Stop: TGA at 325
	SEQ ID NO: 94	92 aa MW at 10382.3kD
NOV20b, CG96637-04 Protein Sequence	MKVSVAALSCLMLVTALGSQARVTKDTETEFMMSKLPLENPVLLDRFHATSADCCIFL TKKGRRFCANPSDKQVQVCMRMLKLDTRIKTRKN	

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 20B.

Table 20B. Comparison of NOV20a against NOV20b.		
Protein Sequence	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV20b	1..116 1..92	72/120 (60%) 78/120 (65%)

Further analysis of the NOV20a protein yielded the following properties shown in Table 20C.

Table 20C. Protein Sequence Properties NOV20a	
PSort analysis:	0.7141 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Cleavage site between residues 22 and 23

A search of the NOV20a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 20D.

Table 20D. Geneseq Results for NOV20a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW05186	Human eosinophil-expressed chemokine protein sequence - Homo sapiens, 137 aa. [WO9632481-A1, 17-OCT-1996]	1..116 1..137	109/137 (79%) 113/137 (81%)	2e-55
AAB62345	Human MPIF-1 splice variant - Homo sapiens, 137 aa. [WO200126676-A1, 19-APR-2001]	1..116 1..137	109/137 (79%) 113/137 (81%)	3e-55
AAW57696	Human MPIF-1 splice variant protein - Homo sapiens, 137 aa. [WO9814582-A1, 09-APR-1998]	1..116 1..137	109/137 (79%) 113/137 (81%)	3e-55
AAU11156	Human G protein-coupled receptor HNFDS78 ligand CKbeta-8 - Homo sapiens, 120 aa. [US6287801-B1, 11-SEP-2001]	1..116 1..120	99/120 (82%) 106/120 (87%)	3e-49
AAW57688	Human MPIF-1 protein - Homo sapiens, 120 aa. [WO9814582-A1, 09-APR-1998]	1..116 1..120	99/120 (82%) 106/120 (87%)	3e-49

In a BLAST search of public sequence databases, the NOV20a protein was found to have homology to the proteins shown in the BLASTP data in Table 20E.

Table 20E. Public BLASTP Results for NOV20a				
Protein Accession Number	Protein/Organism/Length	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P55773	Small inducible cytokine A23 precursor (Macrophage inflammatory protein 3) (MIP-3) (Myeloid progenitor inhibitory factor-1) (MPIF-1) (CK-beta-8) (CKB-8) - Homo sapiens (Human), 120 aa.	1..116 1..120	99/120 (82%) 106/120 (87%)	6e-49
Q16663	Small inducible cytokine A15 precursor (Macrophage inflammatory protein 5) (MIP-5) (Chemokine CC-2) (HCC-2) (NCC-3) (MIP-1 delta) (Leukotactin-1) (LKN-1) (Mrp-2b) - Homo sapiens (Human), 113 aa.	1..106 1..109	64/110 (58%) 80/110 (72%)	5e-27
Q16627	Small inducible cytokine A14 precursor (Chemokine CC-1/CC-3) (HCC- 1/HCC-3) (NCC-2) - Homo sapiens (Human), 93 aa.	1..106 1..91	48/107 (44%) 66/107 (60%)	7e-17
P51670	Small inducible cytokine A9 precursor (Macrophage inflammatory protein 1-gamma) (MIP-1-gamma) (Macrophage inflammatory protein-related protein-2) (MRP-2) (CCF18) - Mus musculus (Mouse), 122 aa.	1..115 1..121	51/124 (41%) 76/124 (61%)	3e-16
P27784	Small inducible cytokine A6 precursor (C10 protein) - Mus musculus (Mouse), 116 aa.	1..110 1..109	45/111 (40%) 66/111 (58%)	2e-15

PFam analysis predicts that the NOV20a protein contains the domains shown in the Table 20F.

Table 20F. Domain Analysis of NOV20a			
Pfam Domain	NOV20a Match Region	Identities/ Similarities for the Matched Region	Expect Value
IL8	40..105	24/70 (34%) 46/70 (66%)	8.9e-17

EXAMPLE 21.

The NOV21 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 21A.

Table 21A. NOV21 Sequence Analysis		
	SEQ ID NO: 95	505 bp
NOV21a, CG97274-01 DNA Sequence	CTCTGTCCCCAGCCCTGCAGCTGCTGCTGTGGCACAGTGCACTCTGGACAGTGCAGGA AGCCACCCCCCTGGGCCCTGCCAGCTCCCTGCCCCAGAGCTTCCTGCTCAAGTGCTTA GAGCAAGTGAGGAAGATCCAGGGCGATGGCGCAGCGCTCCAGGAGAAGCTGGCAGGCT GCTTGAGCCAACTCCATAGCGGCCTTTTCCTCTACCAGGGGCTCCTGCAGGCCCTGGA AGGGATCTCCCCCGAGTTGGGTCCCACCTTGGACACACTGCAGCTGGACGTCGCCGAC TTTGCCACCACCATCTGGCAGCAGATGGAAGAACTGGGAATGGCCCCTGCCCTGCAGC CCACCCAGGGTGCCATGCCGGCCTTCGCCTCTGCTTTCCAGCGCCGGGCAGGAGGGGT CCTGGTTGCCTCCCATCTGCAGAGCTTCCTGGAGGTGTCGTACCGCGTTCTACGCCAC CTTGCCCAGCCCTGAGCCAAGCCCTCCCCATCCCATGTATT	
	ORF Start: at 3	ORF Stop: TGA at 477
	SEQ ID NO: 96	158 aa	MW at 17071.5kD
NOV21a, CG97274-01 Protein Sequence	LSPALQLLLWHSALWTVQEATPLGPASSLPQSFLLKCLEQVRKIQGDGAALQEKLAGC LSQLHSGLFLYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQMEELGMAPALQP TQGAMPAFASAFQRRAGGVLVASHLQSFLEVSYRVLRHLAQP	
	SEQ ID NO: 97	426 bp
NOV21b, CG97274-03 DNA Sequence	GGATCCACCCCCCTGGGCCCTGCCAGCTCCCTGCCCCAGAGCTTCCTGCTCAAGTGCT TAGAGCAAGTGAGGAAGATCCAGGGCGATGGCGCAGCGCTCCAGGAGAAGCTGGCAGG CTGCTTGAGCCAACTCCATAGCGGCCTTTTCCTCTACCAGGGGCTCCTGCAGGCCCTG GAAGGGATCTCCCCCGAGTTGGGTCCCACCTTGGACACACTGCAGCTGGACGTCGCCG ACTTTGCCACCACCATCTGGCAGCAGATGGAAGAACTGGGAATGGCCCCTGCCCTGCA GCCCACCCAGGGTGCCATGCCGGCCTTCGCCTCTGCTTTCCAGCGCCGGGCAGGAGGG GTCCTAGTTGCCTCCCATCTGCAGAGCTTCCTGGAGGTGTCGTACCGCGTTCTACGCC ACCTTGCCCAGCCCCTCGAG	
	ORF Start: at 7	ORF Stop: at 418
	SEQ ID NO: 98	137 aa	MW at 14715.8kD
NOV21b, CG97274-03 Protein Sequence	TPLGPASSLPQSFLLKCLEQVRKIQGDGAALQEKLAGCLSQLHSGLFLYQGLLQALEG ISPELGPTLDTLQLDVADFATTIWQQMEELGMAPALQPTQGAMPAFASAFQRRAGGVL VASHLQSFLEVSYRVLRHLAQ	
	SEQ ID NO: 99	1672 bp
NOV21c, CG97274-04 DNA Sequence	ACCCATGGCTGGACCTGCCACCCAGAGCCCCATGAAGCTGATGGGTGAGTGTCTTGGC CCAGGATGGGAGAGCCGCCTGCCCTGGCATGGGAGGGAGGCTGGTGTGACAGAGGGGC TGGGGATCCCCGTTCTGGGAATGGGGATTAAAGGCACCCAGTGTCCCCGAGAGGGCCT CAGGTGGTAGGGAACAGCATGTCTCCTGAGCCCGCTCTGTCCCCAGCCCTGCAGCTGC TGCTGTGGCACAGTGCACTCTGGACAGTGCAGGAAGCCACCCCCCTGGGCCCTGCCAG CTCCCTGCCCCAGAGCTTCCTGCTCAAGTGCTTAGAGCAAGTGAGGAAGATCCAGGGC GATGGCGCAGCGCTCCAGGAGAAGCTGGTGAGTGAGTGTGCCACCTACAAGCTGTGCC ACCCCGAGGAGCTGGTGCTGCTCGGACACTCTCTGGGCATCCCCTGGGCTCCCCTGAG CAGCTGCCCCAGCCAGGCCCTGCAGCTGGCAGGCTGCTTGAGCCAACTCCATAGCGGC CTTTTCCTCTACCAGGGGCTCCTGCAGGCCCTGGAAGGGATCTCCCCCGAGTTGGGTC CCACCTTGGACACACTGCAGCTGGACGTCGCCGACTTTGCCACCACCATCTGGCAGCA GATGGAAGAACTGGGAATGGCCCCTGCCCTGCAGCCCACCCAGGGTGCCATGCCGGCC TTCGCCTCTGCTTTCCAGCGCCGGGCAGGAGGGGTCCTGGTTGCCTCCCATCTGCAGA	

	GCTTCCTGGAGGTGTCGTACCGCGTTCTACGCCACCTTGCCCAGCCC**TGA**GCCAAGCC CTCCCCATCCCATGTATTTATCTCTATTTAATATTTATGTCTATTTAAGCCTCATATT TAAAGACAGGGAAGAGCAGAACGGAGCCCCAGGCCTCTGTGTCCTTCCCTGCATTTCT GAGTTTCATTCTCCTGCCTGTAGCAGTGAGAAAAAGCTCCTGTCCTCCCATCCCCTGG ACTGGGAGGTAGATAGGTAAATACCAAGTATTTATTACTATGACTGCTCCCCAGCCCT GGCTCTGCAATGGGCACTGGGATGAGCCGCTGTGAGCCCCTGGTCCTGAGGGTCCCCA CCTGGGACCCTTGAGAGTATCAGGTCTCCCACGTGGGAGACAAGAAATCCCTGTTTAA TATTTAAACAGCAGTGTTCCCCATCTGGGTCCTTGCACCCCTCACTCTGGCCTCAGCC GACTGCACAGCGGCCCCTGCATCCCCTTGGCTGTGAGGCCCCTGGACAAGCAGAGGTG GCCAGAGCTGGGAGGCATGGCCCTGGGGTCCCACGAATTTGCTGGGGAATCTCGTTTT TCTTCTTAAGACTTTTGGGACATGGTTTGACTCCCGAACATCACCGACGTGTCTCCTG TTTTTCTGGGTGGCCTCGGGACACCTGCCCTGCCCCCACGAGGGTCAGGACTGTGACT CTTTTTAGGGCCAGGCAGGTGCCTGGACATTTGCCTTGCTGGACGGGGACTGGGGATG TGGGAGGGAGCAGACAGGAGGAATCATGTCAGGCCTGTGTGTGAAAGGAAGCTCCACT GTCACCCTCCACCTCTTCACCCCCCACTCACCAGTGTCCCCTCCACTGTCACATTGTA ACTGAACTTCAGGATAATAAAGTGTTTGCCTCCAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 193	ORF Stop: TGA at 802
	SEQ ID NO: 100	203 aa	MW at 21858.0kD
NOV21c, CG97274-04 Protein Sequence	MSPEPALSPALQLLLWHSALWTVQEATPLGPASSLPQSFLLKCLEQVRKIQGDGAALQ EKLVSECATYKLCHPEELVLLGHSLGIPWAPLSSCPSQALQLAGCLSQLHSGLFLYQG LLQALEGISPELGPTLDTLQLDVADFATTIWQQMEELGMAPALQPTQGAMPAFASAFQ RRAGGVLVASHLQSFLEVSYRVLRHLAQP		
	SEQ ID NO: 101	426 bp
NOV21d, 197208289 DNA Sequence	**GGA**TCCACCCCCCTGGGCCCTGCCAGCTCCCTGCCCCAGAGCTTCCTGCTCAAGTGCT TAGAGCAAGTGAGGAAGATCCAGGGCGATGGCGCAGCGCTCCAGGAGAAGCTGGCAGG CTGCTTGAGCCAACTCCATAGCGGCCTTTTCCTCTACCAGGGGCTCCTGCAGGCCCTG GAAGGGATCTCCCCCGAGTTGGGTCCCACCTTGGACACACTGCAGCTGGACGTCGCCG ACTTTGCCACCACCATCTGGCAGCAGATGGAAGAACTGGGAATGGCCCCTGCCCTGCA GCCCACCCAGGGTGCCATGCCGGCCTTCGCCTCTGCTTTCCAGCGCCGGGCAGGAGGG GTCCTAGTTGCCTCCCATCTGCAGAGCTTCCTGGAGGTGTCGTACCGCGTTCTACGCC ACCTTGCCCAGCCCCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence
	SEQ ID NO: 102	142 aa	MW at 15199.3kD
NOV21d, 197208289 Protein Sequence	GSTPLGPASSLPQSFLLKCLEQVRKIQGDGAALQEKLAGCLSQLHSGLFLYQGLLQAL EGISPELGPTLDTLQLDVADFATTIWQQMEELGMAPALQPTQGAMPAFASAFQRRAGG VLVASHLQSFLEVSYRVLRHLAQPLE		

Sequence comparison of the above protein sequences yields the following
sequence relationships shown in Table 21B.

Table 21B. Comparison of NOV21a against NOV21b through NOV21d.		
Protein Sequence	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV21b	21..157 1..137	137/137 (100%) 137/137 (100%)
NOV21c	1..158 7..203	158/197 (80%) 158/197 (80%)
NOV21d	20..158 2..140	138/139 (99%) 139/139 (99%)

Further analysis of the NOV21a protein yielded the following properties shown in
Table 21C.

Table 21C. Protein Sequence Properties NOV21a	
PSort analysis:	0.5567 probability located in microbody (peroxisome); 0.4273 probability located in mitochondrial matrix space; 0.3175 probability located in lysosome (lumen); 0.1052 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 21 and 22

5

A search of the NOV21a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 21D.

Table 21D. Geneseq Results for NOV21a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW78103	Chimeric receptor agonist polypeptide pMON35783.pep - Homo sapiens, 347 aa. [WO9817810-A2, 30-APR-1998]	21..158 103..275	137/173 (79%) 137/173 (79%)	3e-67
AAR75332	Human granulocyte-colony stimulating factor (G-CSF) - Homo sapiens, 174 aa. [WO9513393-A, 18-MAY-1995]	21..158 1..174	136/174 (78%) 136/174 (78%)	1e-66
AAR15213	[Ser17,27,60,65]huG-CSF - Synthetic, 174 aa. [EP459630-A, 04-DEC-1991]	21..158 1..174	136/174 (78%) 136/174 (78%)	2e-66
AAR15209	[Arg11,40,Ser17,27,60,65]huG-CSF - Synthetic, 174 aa. [EP459630-A, 04-DEC-1991]	21..158 1..174	135/174 (77%) 136/174 (77%)	6e-66
AAR15211	[Ala1,Thr3,Tyr4,Arg5,11,Ser17,27,60,65] huG-CSF - Synthetic, 174 aa. [EP459630-A, 04-DEC-1991]	22..158 2..174	131/173 (75%) 132/173 (75%)	1e-62

5

In a BLAST search of public sequence databases, the NOV21a protein was found to have homology to the proteins shown in the BLASTP data in Table 21E.

Protein Accession Number	Protein/Organism/Length	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
FQHUGL	granulocyte colony-stimulating factor precursor - human, 204 aa.	4..158 14..204	155/191 (81%) 155/191 (81%)	2e-78
CAB58670	SEQUENCE 2 FROM PATENT WO9315211 PRECURSOR - unidentified, 787 aa.	5..158 7..198	142/192 (73%) 146/192 (75%)	5e-68
CAB58682	FRAGMENT C-TER DE LA CHIMERE SAH-G.CSF - unidentified, 177 aa (fragment).	21..158 4..177	138/174 (79%) 138/174 (79%)	6e-68
CAB58669	SEQUENCE 1 FROM PATENT WO9315211 - unidentified, 783 aa.	21..158 610..783	138/174 (79%) 138/174 (79%)	6e-68
E977866	G-CSF PROTEIN - vectors, 177 aa.	21..158 1..177	138/177 (77%) 138/177 (77%)	1e-67

Table 21E. Public BLASTP Results for NOV21a

PFam analysis predicts that the NOV21a protein contains the domains shown in the Table 21F.

Pfam Domain	NOV21a Match Region	Identities/ Similarities for the Matched Region	Expect Value
IL6	41..55	7/15 (47%) 15/15 (100%)	0.017
IL6	56..153	37/102 (36%) 96/102 (94%)	7.3e-49

Table 21F. Domain Analysis of NOV21a

5

EXAMPLE 22.

The NOV22 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 22A.

Table 22A. NOV22 Sequence Analysis		
	SEQ ID NO: 103	399 bp
NOV22a, CG97288-01 DNA Sequence	ATGAGCCTCAGACTTGATACCACCCCTTCCTGTAACAGTGCGAGACCACTTCATGCCT TGCAGGTGCTGCTGCTTCTGTCATTGCTGCTGACTGCTCTGGCTTCCTCCACCAAAGG ACAAACTAAGAGAAACTTGGCGAAAGGCAAAGAGGAAAGTCTAGACAGTGACTTGTAT GCTGAACTCCGCTGCATGTGTATAAAGACAACCTCTGGGAATTCATCCCAAAAACATC CAAAGTTTGGGAAGTCCGGGGAAAGGGAACCCATTGGCAACCAAGTCGAAGTGATAGG CCACACTGAAGGATGGGGAGGGAAATTCTGCCTGGGGACCCCAGATGGTTCCCCAGGA TTCAAGAAAATTGTACAGAAAAAATTGGCAGGTGATGAATCTGCTGATTAA	
	ORF Start: ATG at 1	ORF Stop: TAA at 397
	SEQ ID NO: 104	132 aa	MW at 14128.0kD
NOV22a, CG97288-01 Protein Sequence	MSLRLDTTPSCNSARPLHALQVLLLLSLLLTALASSTKGQTKRNLAKGKEESLDSDLY AELRCMCIKTTSGNSSQKHPKFGKSGEREPIGNQVEVIGHTEGWGGKFCLGTPDGSPG FKKIVQKKLAGDESAD	
	SEQ ID NO: 105	249 bp
NOV22b, CG97288-02 DNA Sequence	ATGAGCCTCAGACTTGATACCACCCCTTCCTGTAACAGTGCGAGACCACTTCATGCCT TGCAGGTGCTGCTGCTTCTGTCATTGCTGCTGACTGCTCTGGCTTCCTCCACCAAAGG ACAAACTAAGAGAAACTTGGCGAAAGGCAAAGAGGAAAGTCTAGACAGTGACTTGTAT GCTGAACTCCGCTGCATGTGTATAAAGACAACCTCTGGGAATTCATCCCAAAAACATC CAAAGTTTGGGAAGTGA	
	ORF Start: ATG at 1	ORF Stop: TGA at 247
	SEQ ID NO: 106	82 aa	MW at 8918.2kD
NOV22b, CG97288-02 Protein Sequence	MSLRLDTTPSCNSARPLHALQVLLLLSLLLTALASSTKGQTKRNLAKGKEESLDSDLY AELRCMCIKTTSGNSSQKHPKFGK	

Sequence comparison of the above protein sequences yields the following

sequence relationships shown in Table 22B.

Table 22B. Comparison of NOV22a against NOV22b.		
Protein Sequence	NOV22a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV22b	1..82 1..82	62/82 (75%) 62/82 (75%)

Further analysis of the NOV22a protein yielded the following properties shown in

Table 22C.

Table 22C. Protein Sequence Properties NOV22a	
PSort analysis:	0.6000 probability located in endoplasmic reticulum (membrane); 0.5994 probability located in mitochondrial inner membrane; 0.3647 probability located in mitochondrial intermembrane space; 0.1802 probability located in mitochondrial matrix space
SignalP analysis:	Cleavage site between residues 35 and 36

A search of the NOV22a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 22D.

Table 22D. Geneseq Results for NOV22a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV22a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB15804	Human chemokine PF4 SEQ ID NO: 46 - Homo sapiens, 128 aa. [WO200042071-A2, 20-JUL-2000]	1..132 1..128	106/137 (77%) 109/137 (79%)	4e-43
AAW96716	A platelet basic protein (PBP) - Homo sapiens, 128 aa. [US5871723-A, 16-FEB-1999]	1..132 1..128	106/137 (77%) 109/137 (79%)	4e-43
AAR13519	Leukocyte derived growth factor - Homo sapiens, 128 aa. [WO9111515-A, 08-AUG-1991]	1..132 1..128	106/137 (77%) 109/137 (79%)	4e-43
AAR05767	Precursor of platelet basic protein (PBP) - Synthetic, 128 aa. [WO9006321-A, 14-JUN-1990]	1..132 1..128	106/137 (77%) 109/137 (79%)	4e-43
AAR13520	Leukocyte derived growth factor analogue - Homo sapiens, 128 aa. [WO9111515-A, 08-AUG-1991]	1..132 1..128	104/137 (75%) 109/137 (78%)	2e-42

In a BLAST search of public sequence databases, the NOV22a protein was found
5   to have homology to the proteins shown in the BLASTP data in Table 22E.

192

Table 22E. Public BLASTP Results for NOV22a				
Protein Accession Number	Protein/Organism/Length	NOV22a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P02775	Platelet basic protein precursor (PBP) [Contains: Connective-tissue activating peptide III (CTAP-III); Low-affinity platelet factor IV (LA-PF4); Beta-thromboglobulin (Beta-TG); Neutrophil-activating peptide 2 (NAP-2)] - Homo sapiens (Human), 128 aa.	1..132 1..128	106/137 (77%) 109/137 (79%)	1e-42
CAC41217	SEQUENCE 11 FROM PATENT WO0136635 - Homo sapiens (Human), 127 aa.	1..130 1..125	75/131 (57%) 87/131 (66%)	5e-27
AAA72500	CONNECTIVE TISSUE ACTIVATING PEPTIDE-III - synthetic construct, 91 aa (fragment).	42..132 5..91	65/96 (67%) 68/96 (70%)	1e-19
AAA73218	COL-CTAP-III(LEU21) FUSION PROTEIN - synthetic construct, 591 aa.	22..132 476..591	69/125 (55%) 78/125 (62%)	1e-18
AAA73216	CTAP-III(LEU21)--HIRUDIN FUSION PROTEIN PRECURSOR - synthetic construct, 162 aa.	44..132 7..91	62/94 (65%) 66/94 (69%)	2e-18

PFam analysis predicts that the NOV22a protein contains the domains shown in the Table 22F.

Table 22F. Domain Analysis of NOV22a			
Pfam Domain	NOV22a Match Region	Identities/ Similarities for the Matched Region	Expect Value
IL8	54..71	9/19 (47%) 18/19 (95%)	0.055

EXAMPLE 23.

The NOV23 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 23A.

Table 23A. NOV23 Sequence Analysis		
	SEQ ID NO: 107	463 bp
NOV23a, CG97516-01 DNA Sequence	GCAGTCCCTCCAGAGACATGGATCCCCAGACAGCACCTTCCCGGGCGCTCCTGCTCCT GCTCTTCTTGCATCTGGCTTTCCTGGGAGGTCGTTCCCACCCGCTGGGCAGCCCCGGT TCAGCCTCGGACTTGGAAACGTCCGGGTTACAGGGAGCAGCGCAACCATTTGCAGGGC AAACTGTCGGAGCTGCAGGTGTCTGGAAGTCCCGGGAGGTAGCCACCGAGGGCATCCG TGGGCACCGCAAAATGGTCCTCTACACCCTGCGGGCACCACGAAGCCCCAAGATGGTG CAAGGGTCTGGCTGCTTTGGGAGGAAGATGGACCGGATCAGCTCCTCCAGTGGCCTGG GCTGCAAAGTGCTGAGGCGGCATTAAGAGGGAGTCCTGGCTGCAGACACCTGCTTCTG ATTCCACAAGGGACTTTTTCCTCAACCCTGTGGCCGCCTTTGAAGTGACTCATTTTT	
	ORF Start: ATG at 18	ORF Stop: TAA at 372
	SEQ ID NO: 108	118 aa	MW at 12498.3kD
NOV23a, CG97516-01 Protein Sequence	MDPQTAPSRALLLLLFLHLAFLGGRSHPLGSPGSASDLETSGLQGAAQPFAGQTVGAA GVWKSREVATEGIRGHRKMVLYTLRAPRSPKMVQGSGCFGRKMDRISSSSGLGCKVLR RH	

Further analysis of the NOV23a protein yielded the following properties shown in

5   Table 23B.

Table 23B. Protein Sequence Properties NOV23a	
PSort analysis:	0.5804 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Cleavage site between residues 27 and 28

A search of the NOV23a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 23C.

Table 23C. Geneseq Results for NOV23a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV23a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB45735	Human BNP prepropeptide - Homo sapiens, 134 aa. [WO200071576-A2, 30-NOV-2000]	1..118 1..134	105/134 (78%) 107/134 (79%)	8e-52
AAY05325	Human gamma-BNP protein sequence - Homo sapiens, 134 aa. [WO9913331-A1, 18-MAR-1999]	1..118 1..134	105/134 (78%) 107/134 (79%)	8e-52
AAR06603	Human Brain Natriuretic Polypeptide - Homo sapiens, 134 aa. [EP385476-A, 05-SEP-1990]	1..118 1..134	105/134 (78%) 107/134 (79%)	8e-52
AAR04087	Protein encoded by human natriuretic related peptide - Sus scrofa, 134 aa. [WO8912069-A, 14-DEC-1989]	1..118 1..134	105/134 (78%) 107/134 (79%)	8e-52
AAB45738	Human BNP propeptide - Homo sapiens, 109 aa. [WO200071576-A2, 30-NOV-2000]	26..118 1..109	80/109 (73%) 82/109 (74%)	3e-37

5

In a BLAST search of public sequence databases, the NOV23a protein was found to have homology to the proteins shown in the BLASTP data in Table 23D.

	Table 23D. Public BLASTP Results for NOV23a			
Protein Accession Number	Protein/Organism/Length	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P16860	Brain natriuretic peptide precursor (BNP) - Homo sapiens (Human), 134 aa.	1..118 1..134	105/134 (78%) 107/134 (79%)	2e-51
Q9N2E7	NATRIURETIC PROTEIN - Gorilla gorilla (gorilla), 134 aa.	1..59 1..59	59/59 (100%) 59/59 (100%)	8e-27
Q9P2Q7	NATRIURETIC PROTEIN - Homo sapiens (Human), 135 aa (fragment).	1..59 1..59	59/59 (100%) 59/59 (100%)	8e-27
Q9N2E8	NATRIURETIC PROTEIN - Pan troglodytes (Chimpanzee), 135 aa (fragment).	1..59 1..59	58/59 (98%) 58/59 (98%)	4e-26
Q9N2E6	NATRIURETIC PROTEIN - Pongo pygmaeus (Orangutan), 135 aa (fragment).	1..59 1..59	54/59 (91%) 56/59 (94%)	9e-24

PFam analysis predicts that the NOV23a protein contains the domains shown in the Table 23E.

	Table 23E. Domain Analysis of NOV23a		
Pfam Domain	NOV23a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ANP	59..116	30/66 (45%) 55/66 (83%)	3.9e-26

5

<u>EXAMPLE 24.</u>

The NOV24 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 24A.

Table 24A. NOV24 Sequence Analysis		
	SEQ ID NO: 109	1583 bp
NOV24a, CG97550-01 DNA Sequence	GTCCCTGCGCTCCCTGCGCCCTGGGGATGCCCCTGCCGCCCTGACGCCCGCCAGCCTG AGCCACCGGCGCATGTGACCGCGCGTCCGCCCCAGTCCCATCCGTAGCGCCCGGCGCC CGGCCCCGCAGCGGCCTCGTTGTNCCCGCCGGCCCCCGCCCGGTCTCCCGCGCTGCCA CCCGCCGCCGGCCCTGCCGCC**ATG**CAGGCGCGAGCGCTGCTCCTGGCCGCGTTGGCCG CGCTGGCGCTGGCCCGGGAGCCCCCTGCGGCGCCGTGTCCCGCGCGCTGCGACGTGTC GCGGTGTCCCAGCCCCCGCTGCCCCGGCGGCTACGTGCCCGACCTCTGCAACTGCTGC CTGGTGTGCGCCGCCAGCGAGGGCGAGCCCTGTGGCGGCCCTCTGGACTCGCCTTGCG GCGAGAGCCTGGAGTGCGTGCGCGGCCTATGCCGCTGCCGCTGGTCGCACGCCGTGTG TGGCACCGACGGGCACACCTATGCCAACGTGTGCGCGCTGCAGGCGGCCAGCCGCCGC GCGCTGCAGCTCTCCGGGACGCCCGTGCGCCAGCTGCAGAAGGGCGCCTGCCCGTTGG GTCTCCACCAGCTGAGCAGCCCGCGCTACAAGTTCAACTTCATTGCTGACGTGGTGGA GAAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTTGGC CGCAACGTGCCCCTGTCCAGCGGTTCTGGCTTCATCATGTCAGAGGCCGGCCTGATCA TCACCAATGCCCACGTGGTGTCCAGCAACAGCGCTGCCCCGGGCAGGCAGCAGCTCAA GGTGCAGCTACAGAATGGGGACTCCTATGAGGCCACCATCAAAGACATCGACAAGAAG TCGGACATTGCCACCATCAAGATCCATCCCAAGAAAAAGCTCCCTGTGTTGTTGCTGG GTCACTCGGCCGACCTGCGGCCTGGGGAGTTTGTGGTGGCCATCGGCAGTCCCTTCGC CCTACAGAACACAGTGACAACGGGCATCGTCAGCACTGCCCAGCGGGAGGGCAGGGAG CTGGGCCTCCGGGACTCCGACATGGACTACATCCAGACGGATGCCATCATCAACTACG GGAACTCCGGGGGACCACTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACGCT CAAGGTCACGGCTGGCATCTCCTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTC ACAGAGTTCCAAGACAAGCAGATCAAAGACTGGAAGAAGCGCTTCATCGGCATACGGA TGCGGACGATCACACCAAGCCTGGTGGATGAGCTGAAGGCCAGCAACCCGGACTTCCC AGAGGTCAGCAGTGGAATTTATGTGCAAGAGGTTGCGCCGAATTCACCTTCTCAGAGA GGCGGCATCCAAGATGGTGACATCATCGTCAAGGTCAACGGGCGTCCTCTAGTGGACT CGAGTGAGCTGCAGGAGGCCGTGCTGACCGAGTCTCCTCTCCTACTGGAGGTGCGGCG GGGGAACGACGACCTCCTCTTCAGCATCGCACCTGAGGTGGTCATGTGAGGGGCGCAT TCCTCCAGCGCCAAGCG
	ORF Start: ATG at 196	ORF Stop: TGA at 1555
	SEQ ID NO: 110	453 aa	MW at 48607.2kD
NOV24a, CG97550-01 Protein Sequence	MQARALLLAALAALALAREPPAAPCPARCDVSRCPSPRCPGGYVPDLCNCCLVCAASE GEPCGGPLDSPCGESLECVRGLCRCRWSHAVCGTDGHTYANVCALQAASRRALQLSGT PVRQLQKGACPLGLHQLSSPRYKFNFIADVVEKIAPAVVHIELFLRHPLFGRNVPLSS GSGFIMSEAGLIITNAHVVSSNSAAPGRQQLKVQLQNGDSYEATIKDIDKKSDIATIK IHPKKKLPVLLLGHSADLRPGEFVVAIGSPFALQNTVTTGIVSTAQREGRELGLRDSD MDYIQTDAIINYGNSGGPLVNLDGEVIGINTLKVTAGISFAIPSDRITRFLTEFQDKQ IKDWKKRFIGIRMRTITPSLVDELKASNPDFPEVSSGIYVQEVAPNSPSQRGGIQDGD IIVKVNGRPLVDSSELQEAVLTESPLLLEVRRGNDDLLFSIAPEVVM

5

Further analysis of the NOV24a protein yielded the following properties shown in Table 24B.

Table 24B. Protein Sequence Properties NOV24a	
PSort analysis:	0.3700 probability located in outside; 0.1080 probability located in nucleus; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 18 and 19

A search of the NOV24a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 24C.

Table 24C. Geneseq Results for NOV24a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE14349	Human protease PRTS-14 protein - Homo sapiens, 453 aa. [WO200183775-A2, 08-NOV-2001]	1..453 1..453	453/453 (100%) 453/453 (100%)	0.0
AAY93961	A HtrA-2 (high temperature requirement A-2) protein - Homo sapiens, 453 aa. [WO200039149-A2, 06-JUL-2000]	1..453 1..453	453/453 (100%) 453/453 (100%)	0.0
AAY93963	A murine HtrA-2 (high temperature requirement A-2) protein - Mus sp, 348 aa. [WO200039149-A2, 06-JUL-2000]	109..453 4..348	345/345 (100%) 345/345 (100%)	0.0
AAU31560	Novel human secreted protein #2051 - Homo sapiens, 351 aa. [WO200179449-A2, 25-OCT-2001]	106..453 1..351	331/351 (94%) 333/351 (94%)	0.0
AAU16943	Human novel secreted protein, SEQ ID 184 - Homo sapiens, 330 aa. [WO200155441-A2, 02-AUG-2001]	55..350 28..323	293/296 (98%) 294/296 (98%)	e-169

In a BLAST search of public sequence databases, the NOV24a protein was found to have homology to the proteins shown in the BLASTP data in Table 24D.

Table 24D. Public BLASTP Results for NOV24a				
Protein Accession Number	Protein/Organism/Length	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9D236	Probable serine protease HTRA3 precursor (EC 3.4.21.-) (Toll-associated serine protease) - Mus musculus (Mouse), 460 aa.	1..453 1..460	408/460 (88%) 424/460 (91%)	0.0
P83110	Probable serine protease HTRA3 precursor (EC 3.4.21.-) - Homo sapiens (Human), 452 aa.	1..453 1..452	412/455 (90%) 416/455 (90%)	0.0
Q92743	Serine protease HTRA1 precursor (EC 3.4.21.-) (L56) - Homo sapiens (Human), 480 aa.	6..452 14..478	270/472 (57%) 343/472 (72%)	e-145
Q9QZK5	INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 5 PROTEASE - Rattus norvegicus (Rat), 480 aa.	6..452 13..478	268/473 (56%) 341/473 (71%)	e-143
Q9QZK6	INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 5 PROTEASE - Mus musculus (Mouse), 480 aa.	5..452 10..478	269/476 (56%) 342/476 (71%)	e-143

PFam analysis predicts that the NOV24a protein contains the domains shown in the Table 24E.

Table 24E. Domain Analysis of NOV24a			
Pfam Domain	NOV24a Match Region	Identities/ Similarities for the Matched Region	Expect Value
IGFBP	25..63	20/53 (38%) 29/53 (55%)	0.041
Kazal	76..126	19/62 (31%) 33/62 (53%)	3.5e-05
Trypsin	170..341	48/247 (19%) 136/247 (55%)	4.3e-16
PDZ	348..439	21/100 (21%) 69/100 (69%)	7.3e-09

EXAMPLE 25.

The NOV25 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 25A.

Table 25A. NOV25 Sequence Analysis		
	SEQ ID NO: 111	402 bp
NOV25a, CG97738-01 DNA Sequence	ATATTTAATTTTAAAAC**TATG**ATGATTGCTTCCTGCTCCTCTCCTCCTCCACGGGCTC CCTGGAGTCCTTGCAAGCTGGCCAGGATGTCTCAGGCTGAGTTTGAGAGAGCTGTGGA AGACGTTAAACACCTTAAGACCAAGCCAGGGGATGATGAGGATGTGTTCCTCTATGGC CACTACAAACAAGCAACTGTGGGCGACATAAATACAGAATGGCCTGGGATGTTGGATT TCAAAGGCAAGACCAAGTGGGATGCCTGGAATGAGCTGAAAGGGACTACCAAGGAAGA TGCCATGAAAGCTTACGTCAACAATGTAGAAGAGCTAAGGAAAAAACATGGAATG**TAA** GAGACTGGATTTGGTTGCCAGCCATGTGTTTATCCTAAACTGAGACAGTGCCTT	
	ORF Start: ATG at 19	ORF Stop: TAA at 346
	SEQ ID NO: 112	109 aa	MW at 12446.1kD
NOV25a, CG97738-01 Protein Sequence	MMIASCSSPPPRAPWSPCKLARMSQAEFERAVEDVKHLKTKPGDDEDVFLYGHYKQAT VGDINTEWPGMLDFKGKTKWDAWNELKGTTKEDAMKAYVNNVEELRKKHGM	

Further analysis of the NOV25a protein yielded the following properties shown in Table 25B.

Table 25B. Protein Sequence Properties NOV25a	
PSort analysis:	0.3600 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV25a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 25C.

Table 25C. Geneseq Results for NOV25a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV25a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAP60958	Sequence of human endogenous benzodiazepineoid (EBZD) polypeptide - Homo sapiens, 107 aa. [WO8604239-A, 31-JUL-1986]	5..109 3..107	82/105 (78%) 94/105 (89%)	3e-46
AAG75960	Human colon cancer antigen protein SEQ ID NO:6724 - Homo sapiens, 106 aa. [WO200122920-A2, 05-APR-2001]	4..109 1..106	80/106 (75%) 93/106 (87%)	3e-45
AAP60957	Sequence of bovine endogenous benzodiazepineoid (EBZD) polypeptide - Bos taurus, 111 aa. [WO8604239-A, 31-JUL-1986]	4..109 6..111	79/106 (74%) 93/106 (87%)	4e-44
AAY92053	HrPCa13 polypeptide, endozepine, from androgen-inducible gene clone - Homo sapiens, 87 aa. [WO200018961-A2, 06-APR-2000]	23..109 1..87	69/87 (79%) 80/87 (91%)	8e-38
AAR11875	Recombinant human EBZD - Homo sapiens, 86 aa. [US5011777-A, 30-APR-1991]	24..109 1..86	68/86 (79%) 79/86 (91%)	3e-37

5

In a BLAST search of public sequence databases, the NOV25a protein was found to have homology to the proteins shown in the BLASTP data in Table 25D.

Table 25D. Public BLASTP Results for NOV25a				
Protein Accession Number	Protein/Organism/Length	NOV25a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAD19062	SEQUENCE 10 FROM PATENT WO0018961 - Homo sapiens (Human), 87 aa.	23..109 1..87	69/87 (79%) 80/87 (91%)	2e-37
CAA44618	ACYL-COA-BINDING PROTEIN /DIAZEPAM-BINDING INHIBITOR - synthetic construct, 87 aa.	23..109 1..87	68/87 (78%) 80/87 (91%)	7e-37
P07108	Acyl-CoA-binding protein (ACBP) (Diazepam binding inhibitor) (DBI) (Endozepine) (EP) - Homo sapiens (Human), 86 aa.	24..109 1..86	68/86 (79%) 79/86 (91%)	7e-37
P07107	Acyl-CoA-binding protein (ACBP) (Diazepam binding inhibitor) (DBI) (Endozepine) (EP) - Bos taurus (Bovine), 86 aa.	24..109 1..86	67/86 (77%) 79/86 (90%)	3e-36
Q9TSG2	ENDOZEPINE - Sus scrofa (Pig), 87 aa.	23..109 1..87	67/87 (77%) 78/87 (89%)	2e-35

PFam analysis predicts that the NOV25a protein contains the domains shown in the Table 25E.

Table 25E. Domain Analysis of NOV25a			
Pfam Domain	NOV25a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ACBP	24..108	52/89 (58%) 75/89 (84%)	1.7e-46

5

<u>EXAMPLE 26.</u>

The NOV26 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 26A.

Table 26A. NOV26 Sequence Analysis		
	SEQ ID NO: 113	918 bp
NOV26a, CG97800-01 DNA Sequence	<u>TATAAGTGTGCCCCAGCCCATCCCGATGGTCAGCCAGTCCTGAGCACCTAACC**ATGTT**</u> <u>**G**GGCATCACTGTCCTCGCTGCGCTCTTGGCCTGTGCCTCCACCTGTGGGGTGCCCAGC</u> TTCCCGCCCAACCTATCCGCCCGAGTGGTGGGAGGAGAGGATGCCCGGCCCCACAGCT GGCCCTGGCAGATCTCCCTCCAGTACCTCAAGAACGACACGTGGAGGCATACGTGTGG CGGGACTTTGATTGCTAGCAACTTCGTCCTCACTGCCGCCCACTGCATCAGCAACACC CGGACCTACCGTGTGGCCGTGGGAAAGAACAACCTGGAGGTGGAAGACGAAGAAGGAT CCCTGTTTGTGGGTGTGGACACCATCCACGTCCACAAGAGATGGAATGCCCTCCTGTT GCGCAATGATATTGCCCTCATCAAGCTTGCAGAGCATGTGGAGCTGAGTGACACCATC CAGGTGGCCTGCCTGCCAGAGAAGGACTCCCTGCTCCCCAAGGACTACCCCTGCTATG TCACCGGCTGGGGCCGCCTCTGGAGGGGACTCCGGTGGCCCACTGAACTGCCAGTTGG AGAACGGTTCCTGGGAGGTGTTTGGCATCGTCAGCTTTGGCTCCCGGCGGGGCTGCAA CACCCGCAAGAAGCCGGTAGTCTACACCCGGGTGTCCGCCTACATCGACTGGATCAAC GAGAAAATGCAGCTGTGATTTGTTGCTGGGAGCGGCGGCAGCGAGTCCCTGCAACAGC AATAAACTTCCTTCTCCTCGGGCCACCTGGATCCTTGATTTGTGCAGCTTCTGTTGCT TCCCTCCTCTCTGGTGCTGCCCCTTTCCACACTATGGAGCCAAAGAGAGACCCCACTC AGCCAGTTTCCCCCACCCTGCAT**TAG**<u>ACAGGTGGGGAAACAGAGGCCG</u>
	ORF Start: ATG at 54	ORF Stop: TAG at 894
	SEQ ID NO: 114	280 aa	MW at 30908.2kD
NOV26a, CG97800-01 Protein Sequence	MLGITVLAALLACASTCGVPSFPPNLSARVVGGEDARPHSWPWQISLQYLKNDTWRHT CGGTLIASNFVLTAAHCISNTRTYRVAVGKNNLEVEDEEGSLFVGVDTIHVHKRWNAL LLRNDIALIKLAEHVELSDTIQVACLPEKDSLLPKDYPCYVTGWGRLWRGLRWPTELP VGERFLGGVWHRQLWLPAGLQHPQEAGSLHPGVRLHRLDQRENAAVICCWERRQRVPA TAINFLLLGPPGSLICAASVASLLSGAAPFHTMEPKRDPTQPVSPTLH		
	SEQ ID NO: 115	812 bp
NOV26b, CG97800-02 DNA Sequence	<u>AGCCAGTCCTGAGCACCTAACCATGTTGGGCATCACTGTCCTCGCTGCGCTCTTGGCC</u> <u>TGTGCCTCCAGCTGTGGGGTGCCCAGCTTCCCGCCCAACCTATCCGCCCGAGTGGTGG</u> <u>GAGGAGAGG**ATG**CCCGGCCCCACAGCTGGCCCTGGCAGCAACACCCGGACCTACCGTG</u> TGGCCGTGGGAAAGAACAACCTGGAGGTGGAAGACGAAGAAGGATCCCTGTTTGTGGG TGTGGACACCATCCACGTCCACAAGAGATGGAATGCCCTCCTGTTGCGCAATGATATT GCCCTCATCAAGCTTGCAGAGCATGTGGAGCTGAGTGACACCATCCAGGTGGCCAGCC TGCCAGAGAAGGACTCCCTGCTCCCCAAGGACTACCCCTGCTATGTCACCGGCTGGGG CCGCCTCTGGAGGGGACTCCGGTGGCCCACTGAACTGCCAGTTGGAGAACGGTTCCTG GGAGGTGTTTGGCATCGTCAGCTTTGGCTCCCGGCGGGGCTGCAACACCCGCAAGAAG CCGGTAGTCTACACCCGGGTGTCCGCCTACATCGACTGGATCAACGAGAAAATGCAGC TGTGATTTGTTGCTGGGAGCGGCGGCAGCGAGTCCCTGCAACAGCAATAAACTTCCTT CTCCTCGGGCCACCTGGATCCTTGATTTGTGCAGCTTCTGTTGCTTCCCTCCTCTCTG GTGCTGCCCCTTTCCACACTATGGAGCCAAAGAGAGACCCCACTCAGCCAGTTTCCCC CACCCTGCATGGT**TAG**GCAGGTGGGAAACAGAGGATGGTTAGGTGATCAGGACTGGCT
	ORF Start: ATG at 126	ORF Stop: TAG at 768
	SEQ ID NO: 116	214 aa	MW at 23625.9kD
NOV26b, CG97800-02 Protein Sequence	MPGPTAGPGSNTRTYRVAVGKNNLEVEDEEGSLFVGVDTIHVHKRWNALLLRNDIALI KLAEHVELSDTIQVASLPEKDSLLPKDYPCYVTGWGRLWRGLRWPTELPVGERFLGGV WHRQLWLPAGLQHPQEAGSLHPGVRLHRLDQRENAAVICCWERRQRVPATAINFLLLG PPGSLICAASVASLLSGAAPFHTMEPKRDPTQPVSPTLHG		
	SEQ ID NO: 117	918 bp

NOV26c, CG97800-03 DNA Sequence	TATAAGTGTGCCCCAGCCCATCCCGATGGTCAGCCAGTCCTGAGCACCTAACCATGTT GGGCATCACTGTCCTCGCTGCGCTCTTGGCCTGTGCCTCCACCTGTGGGGTGCCCAGC TTCCCGCCCAACCTATCCGCCCGAGTGGTGGGAGGAGAGGATGCCCGGCCCCACAGCT GGCCCTGGCAGATCTCCCTCCAGTACCTCAAGAACGACACGTGGAGGCATACGTGTGG CGGGACTTTGATTGCTAGCAACTTCGTCCTCACTGCCGCCCACTGCATCAGCAACACC CGGACCTACCGTGTGGCCGTGGGAAAGAACAACCTGGAGGTGGAAGACGAAGAAGGAT CCCTGTTTGTGGGTGTGGACACCATCCACGTCCACAAGAGATGGAATGCCCTCCTGTT GCGCAATGATATTGCCCTCATCAAGCTTGCAGAGCATGTGGAGCTGAGTGACACCATC CAGGTGGCCTGCCTGCCAGAGAAGGACTCCCTGCTCCCCAAGGACTACCCCTGCTATG TCACCGGCTGGGGCCGCCTCTGGAGGGGACTCCGGTGGCCCACTGAACTGCCAGTTGG AGAACGGTTCCTGGGAGGTGTTTGGCATCGTCAGCTTTGGCTCCCGGCGGGGCTGCAA CACCCGCAAGAAGCCGGTAGTCTACACCCGGGTGTCCGCCTACATCGACTGGATCAAC GAGAAAATGCAGCTGTGATTTGTTGCTGGGAGCGGCGGCAGCGAGTCCCTGCAACAGC AATAAACTTCCTTCTCCTCGGGCCACCTGGATCCTTGATTTGTGCAGCTTCTGTTGCT TCCCTCCTCTCTGGTGCTGCCCCTTTCCACACTATGGAGCCAAAGAGAGACCCCACTC AGCCAGTTTCCCCCACCCTGCATTAGACAGGTGGGGAAACAGAGGCCG
	ORF Start: ATG at 54 . ORF Stop: TAG at 894
	SEQ ID NO: 118 280 aa MW at 30908.2kD
NOV26c, CG97800-03 Protein Sequence	MLGITVLAALLACASTCGVPSFPPNLSARVVGGEDARPHSWPWQISLQYLKNDTWRHT CGGTLIASNFVLTAAHCISNTRTYRVAVGKNNLEVEDEEGSLFVGVDTIHVHKRWNAL LLRNDIALIKLAEHVELSDTIQVACLPEKDSLLPKDYPCYVTGWGRLWRGLRWPTELP VGERFLGGVWHRQLWLPAGLQHPQEAGSLHPGVRLHRLDQRENAAVICCWERRQRVPA TAINFLLLGPPGSLICAASVASLLSGAAPFHTMEPKRDPTQPVSPTLH

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 26B.

Table 26B. Comparison of NOV26a against NOV26b and NOV26c.		
Protein Sequence	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV26b	77..280 10..213	191/204 (93%) 191/204 (93%)
NOV26c	1..280 1..280	268/280 (95%) 268/280 (95%)

Further analysis of the NOV26a protein yielded the following properties shown in Table 26C.

Table 26C. Protein Sequence Properties NOV26a	
PSort analysis:	0.8650 probability located in lysosome (lumen); 0.6854 probability located in outside; 0.1092 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 19 and 20

A search of the NOV26a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 26D.

204

Table 26D. Geneseq Results for NOV26a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV26a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAR90683	Human caldecrin contg. preprosequence - Rattus sp, 268 aa. [WO9600287-A, 04-JAN-1996]	1..183 1..177	165/183 (90%) 170/183 (92%)	2e-94
AAR88481	Human elastase IV protein - Homo sapiens, 268 aa. [WO9601270-A1, 18-JAN-1996]	1..183 1..177	164/183 (89%) 169/183 (91%)	2e-93
AAY51839	Human elastase IV homolog HEIV protein fragment - Homo sapiens, 268 aa. [US6030791-A, 29-FEB-2000]	1..164 1..164	160/164 (97%) 161/164 (97%)	1e-92
AAW89410	Human homologue of rat elastase IV - Homo sapiens, 268 aa. [US5856109-A, 05-JAN-1999]	1..164 1..164	160/164 (97%) 161/164 (97%)	1e-92
AAW40530	Human elastase homologue HEIV protein - Homo sapiens, 268 aa. [US5738991-A, 14-APR-1998]	1..164 1..164	160/164 (97%) 161/164 (97%)	1e-92

In a BLAST search of public sequence databases, the NOV26a protein was found to have homology to the proteins shown in the BLASTP data in Table 26E.

Table 26E. Public BLASTP Results for NOV26a				
Protein Accession Number	Protein/Organism/Length	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q99895	Caldecrin precursor (EC 3.4.21.2) (Chymotrypsin C) - Homo sapiens (Human), 268 aa.	1..183 1..177	166/183 (90%) 171/183 (92%)	9e-95
S68826	pancreatic elastase (EC 3.4.21.36) isoform 2 precursor - human, 268 aa.	1..183 1..177	165/183 (90%) 170/183 (92%)	8e-94
P55091	Caldecrin precursor (EC 3.4.21.2) (Chymotrypsin C) (Serum calcium-decreasing factor) - Rattus norvegicus (Rat), 268 aa.	1..164 1..164	125/164 (76%) 145/164 (88%)	1e-72
JQ1473	pancreatic elastase (EC 3.4.21.36) IV precursor - rat, 268 aa.	1..164 1..164	110/164 (67%) 127/164 (77%)	7e-59
Q9W7Q0	ELASTASE 3 - Paralichthys olivaceus (Flounder), 266 aa.	6..166 5..163	97/161 (60%) 121/161 (74%)	5e-51

PFam analysis predicts that the NOV26a protein contains the domains shown in the Table 26F.

Table 26F. Domain Analysis of NOV26a			
Pfam Domain	NOV26a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Trypsin	30..166	60/159 (38%) 112/159 (70%)	1.2e-47

5

EXAMPLE 27.

The NOV27 clone was analyzed, and the nucleotide and encoded polypeptide

sequences are shown in Table 27A.

Table 27A. NOV27 Sequence Analysis		
	SEQ ID NO: 119	3290 bp
NOV27a, CG98092-01 DNA Sequence	GCCGGGAGGGCCGCGTGAGGAGAGCGAAGAGGGAGCCCGAGCTCTGCGGCCCCGGGTG GCGGGCCGGGGGCGCCCGTGAGCAGAGACCTCCCCGTCGACGGGGGGCGATGTTCCCC TCACCCGTGGGCTCTTCGGGCAGCCAGGGCATCCCGCCGCTGGGACCTGGACCCGCTC AGTGGGCACCCCTCAGGCTGCTCC**ATG**GAGACCCTGTGCCCTGCGCCCCGCCTGGCAG TGCCGGCGTCCCCGCGAGGGTCGCCCTGCTCCCCCACGCCCCGGAAGCCGTGTCGGGG GACCCAGGAATTCTCTCCGCTGTGCCTGCGTGCCCTCGCCTTCTGCGCCCTTGCCAAG CCCCGGGCGTCCTCTCTGGGCCCGGGGCCTGGGGAGCTGGCGGCGCGGTCCCCAGTGC TGCGGGCCCTCAGGCCCCCCTGCGCCCTGGCGGCTGGGCCCCGGATGGCCTGAAGCA CCTCTGGGCACCGACCGGGCGGCCCGGCGTTCCTAACACCGCCGCCGGCGAGGATGCG GACGTCGCAGCGTGCCCCCGCCGCGGAGAGGAGGAAGAGGGCGGAGGCGGTTTCCCGC ACTTCGGCGTTCGCTCCTGTGCACCTCCGGGCCGCTGCCCTGCGCCCCCGCACCCTCG GGAATCTACGACCAGCTTCGCCTCGGCCCCGCCTCGCCCGGCCCCGGGTCTCGAGCCT CAGCGTGGCCCAGCCGCCAGCCCGCCTCAGGAACCCAGTTCCCGGCCTCCGTCGCCAC CTGCGGGCCTCTCCACCGAGCCCGCGGGTCCCGGGACGGCGCCGCGGCCGTTCCTGCC CGGCCAGCCTGCCGAAGTCGATGGAAACCCCCGCCGGCCGCCCCCGAGGCTCCAGCG GCCAGCCCCTCGACGGCCAGCCCGGCTCCGGCCGCACCCGGAGATCTCCGCCAGGAAC ATTTCGATCGTCTGATCCGCCGGTCGAAACTTTGGTGTTACGCGAAGGGCTTCGCCTT GGACACTCCGAGTTTGCGCCGGGGGCCAGAGCGGCCGCCTGCGAAAGGGCCGGCTCGG GGAGCCGCCAAGAAACGCCGGCTGCCGGCGCCCCCTCCGCGCACCGCGCAGCCCCGCC GCCCTGCACCGACGCTCCCCACCACGAGCACCTTCAGCCTCCTCAACTGCTTCCCCTG CCCCCCGGCCCTGGTGGTGGGGGAAGACGGAGACCTAAAGCCGGCATCCTCGCTTCGC CTCCAGGGAGACTCTAAGCCCCCGCCCGCCCACCCGCTGTGGAGGTGGCAGATGGGGG GTCCCGCTGTCCCTGAGCCCCCTGGCCTCAAATTCTGGGGGATCAACATGGATGAAAG **CTGA**CCGTGGGACTTCTGCCAAAGGGGAAAAGTTGGGACCATGGCCAAACCGCGGGCT TGAGGAGGGAGCCCGTTTCTCACATTTGTCCCCTTCCTTTACATTTTAGGAGCTGTG GGCAGAGGGACCTAAATAACAGTGATCTTCATTCAAGCACCTAAGTTTTCGGGGTGAC AGTCCCTCCCCCTCATCCTTTGCAGAGGAACCCAGGGCTGGAGTCGGGAGAAGGCTGA TGACATAGATTCCAATCCCTGCCTCCTTCCATCTCGGACCGTTGGAGGCAGGGCCTGC ACCCCAGTGGGAGCAAAGGAGGCCACCGCTCAAAGACACCCCCCCACCCAAAAAAAAG GGGAGAAGAGAGAGACCTCGGTGATGGACAAACCGGTTGTTACTGTGTCTGTGGGCGA GCCTGGGGTGCGGGGCTGTGGTGGGGGTGGGGAGATGATTGGCAGCTCCCTGGGGGCA TCCCCCACCCCCACTGTCCAGGCCTTTAACCCTTTGCTCCCCTCAGGCCTTCCCTAAC GCTCCAAGCACCGCTGGAGCCTTTAATGGGTGAGGGAACTTGGGTAAGAGGAAGATCA CCCCCTTCCTGTCCCCTTTCTAGGCCCCCTCAAGTGCAGGTGACCCTTAATTGGTGAG ATCTTCAGCCTCAGCCGCCGACCTTTCCCTTTTGTCCAGTTTTGGAGTTCCCGTTTTT TCCTTGTTTGCTTTCCGAGTGTAAGGTCTGGCCGGTGAGAAAGATTTCCCCCAACCTT GATTAATCAGCCCCCTCCCCCAACTTACTTCCCTTAGGACGGGTAGGGCTGAGGGACC TCCTCTCCTGGAAAGTGCTTACTTTGCCTGGGGAAGGGGCTAGACACTGTCCCAGGGA AAGTAATAGAAGGTGGAAGAAATCAATAAAATCAGACCAAACAAGTCGCCTTTCGAGG GCCTCCACCGATTTATGGATGAGAGGGGGTGGAGGTGGAAGGCAGGCCCAAGTCCATT CTTTGGACACCCAAACTCAGCCCCCTTAAAGAGTGGAAACAAAACAAGCTGCACTTTG CAGAGGTGGTAAATGAAAGGACTCTTGGCCTAACTTCAAGAGTCCCCTGGGGTTTGAA GGGGCAAAGTTTGAGTCTGGATGGAACCTGGGCTGAGGTACCTTAAGCTTCCCCCCGC AACACCCCAGCCTCAGGGATTGCGGGAGTTGTCAGAGATCTGATGGATCCGAAAGGGG CAGGGCCAGGGATTAGGTTTGGGGTCAGAGGTTCTGTTTTCCAGGGGAGGGGTGAGA TAGGCCTGGATCATGCCCTCTGCCATGCCCTCCAGCTAGGAGGATCTTGAGTCAGAGA GGATTGGAAGTGCTTTCTCCTCCACCCAGGTGAGGTCAGGGGAGCTTAGGTCTTAGGG AGATGGCAAGTTGAGGTATGAAGGGAAGCTGGGCTTTTGGAGCTGCCGAACAACTGA GGGACCCAGTGCGCCTTCCATCCCGCACTAGTGAATAGCGCCCCTCTTCCCCCGAAA ACGAGGTGCGAGAGGAACAATTCCCACGCTGGGGAAGGACTTGTCTCCTTTTCTGTGA

207

	AAATGCTTTGTAAAAAGTTGTTATTGTTTGCATAGAGCAGATTCTTGAGAAAAACTGT TTTGGACCATAAAAGTTTTGTTTGTTTTAAAAACTGTCTCCTTTCATTTTTCTTTCCT TGGGGGTGAGGGTGGGGGTGGGGTGGGTGGGCACTTCTCTCTTTTTCCTTAACATCTG GCCTCTGTGAACCCTGCTGACCTCCCCTCCCCCTCCAGCTGTGTTGTGGGAGGAGGAG GAAGAAGGGGTGGGGGAGTGCCTTCCACCCTGTGCTTCGGGAGTCTCCATCTTATTTT GCCCCCCAAGAATGGAGAACGGGAGGAAGAAGACACAAGGGGTGGGGAGAGAATCGCT GTGAAGAGGGGGGCTGTCAGAAGTCTGGAAAGGCAGACTCCC
	ORF Start: ATG at 199	ORF Stop: TGA at 1336
	SEQ ID NO: 120	379 aa	MW at 39321.2kD
NOV27a, CG98092-01 Protein Sequence	METLCPAPRLAVPASPRGSPCSPTPRKPCRGTQEFSPLCLRALAFCALAKPRASSLGP GPGELAARSPVLRGPQAPLRPGGWAPDGLKHLWAPTGRPGVPNTAAGEDADVAACPRR GEEEEGGGGFPHFGVRSCAPPGRCPAPPHPRESTTSFASAPPRPAPGLEPQRGPAASP PQEPSSRPPSPPAGLSTEPAGPGTAPRPFLPGQPAEVDGNPPPAAPEAPAASPSTASP APAAPGDLRQEHFDRLIRRSKLWCYAKGFALDTPSLRRGPERPPAKGPARGAAKKRRL PAPPPRTAQPRRPAPTLPTTSTFSLLNCFPCPPALVVGEDGDLKPASSLRLQGDSKPP PAHPLWRWQMGGPAVPEPPGLKFWGINMDES

Further analysis of the NOV27a protein yielded the following properties shown in Table 27B.

Table 27B. Protein Sequence Properties NOV27a	
PSort analysis:	0.3000 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.2584 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	Cleavage site between residues 50 and 51

A search of the NOV27a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 27C.

Table 27C. Geneseq Results for NOV27a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV27a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW31855	Mycobacterium tuberculosis 55 kDa protein - Mycobacterium tuberculosis, 572 aa. [WO9741252-A2, 06-NOV-1997]	6..368 223..537	102/378 (26%) 126/378 (32%)	2e-14
AAW31852	Mycobacterium tuberculosis 74 kDa protein - Mycobacterium tuberculosis, 763 aa. [WO9741252-A2, 06-NOV-1997]	6..368 414..728	102/378 (26%) 126/378 (32%)	2e-14
ABB70063	Drosophila melanogaster polypeptide SEQ ID NO 36981 - Drosophila melanogaster, 446 aa. [WO200171042-A2, 27-SEP-2001]	13..368 103..398	97/372 (26%) 119/372 (31%)	6e-14
AAW72204	HSV-2 strain SB5 Contig ID 15 ORF#39 protein - Herpes simplex virus type 2, 3119 aa. [WO9820016-A1, 14-MAY-1998]	12..350 2627..2993	107/385 (27%) 128/385 (32%)	4e-13
ABG21919	Novel human diagnostic protein #21910 - Homo sapiens, 325 aa. [WO200175067-A2, 11-OCT-2001]	37..237 3..186	60/204 (29%) 76/204 (36%)	1e-12

5

In a BLAST search of public sequence databases, the NOV27a protein was found to have homology to the proteins shown in the BLASTP data in Table 27D.

Table 27D. Public BLASTP Results for NOV27a				
Protein Accession Number	Protein/Organism/Length	NOV27a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q99JK6	HYPOTHETICAL 33.7 KDA PROTEIN - Mus musculus (Mouse), 327 aa (fragment).	47..378 5..326	239/332 (71%) 261/332 (77%)	e-136
Q9FPQ6	Vegetative cell wall protein gp1 precursor (Hydroxyproline-rich glycoprotein 1) - Chlamydomonas reinhardtii, 555 aa.	13..352 43..341	97/352 (27%) 117/352 (32%)	4e-18
Q95JD0	BASIC PROLINE-RICH PROTEIN - Sus scrofa (Pig), 511 aa.	6..370 180..490	111/386 (28%) 120/386 (30%)	2e-17
Q95JD1	BASIC PROLINE-RICH PROTEIN - Sus scrofa (Pig), 566 aa.	6..308 309..564	93/323 (28%) 99/323 (29%)	4e-17
Q95JC9	BASIC PROLINE-RICH PROTEIN - Sus scrofa (Pig), 676 aa.	6..370 130..463	106/379 (27%) 114/379 (29%)	4e-16

PFam analysis predicts that the NOV27a protein contains the domains shown in the Table 27E.

Table 27E. Domain Analysis of NOV27a			
Pfam Domain	NOV27a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

EXAMPLE 28.

The NOV28 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 28A.

Table 28A. NOV28 Sequence Analysis		
	SEQ ID NO: 121	1520 bp
NOV28a, CG98121-01 DNA Sequence	GGCAGCGGGGCAGGCGCGTGGCCGGGCCGCGGCGCGATGAGTGGGGCCCGGGCGGCGC CCGGGGCCGCGGGCAACGGCGCGGTCCGGGGGCTGCGGGTGGACGGGCTGCCCCCGCT GCCAAAGAGCTTGAGCGGGCTGCTGCACTCGGCGTCGGGCGGCGGCGCGTCTGGGGGC TGGCGGCACCTGGAGCGGCTGTACGCGCAGAAGTCGCGCATCCAGGACGAGCTGAGCC GCGGGGGCCCGGGCGGCGGCGGGGCCCGGGCGGGCAGCGCTGCCCGCCAAGCCTCCCA ACCTGGACGCCGCTCTGGCGCTGCTCCGCAAAGAGATGGTTGGTCTCCGCCAGCTGGA CATGTCCTTGCTCTGCCAACTGTACAGCCTCTACGAGTCGATTCAGGAGTACAAGGGG GCATGCCAGGCAGCCTCCAGCCCAGACTGCACTTACGCTCTGGAGAACGGCTTCTTCG **ATGA**AGAGGAGGAATATTTCCAGGAGCAGAACTCCCTGCACGACAGGAGGGACCGAGG CCCTCCTCGGGACTTGTCACTGCCTGTCTCCTCCCTCTCCAGCAGCGACTGGATTCTG GAGTCCATCTAGAGGGTCTTGGGAGGGATGTGACTGTTGGGAAGCCCTTCCTACTGGA CACGCTGTCATCATTTGCTGCTTCTCTTGCAAGAAAGCACCTCCGTTGTGGACGGTCC TCGGGCACAGGGGATGAGCGCTACCAGTTTCATTTGTAGGCAGGGAGTTCTCCGCGGA TGCATGGTGGCAGTCTGCTTTGATGGCAGCAGTTTCTGCTTAGGTGACCTAGAGGTCC TCAGCAGTATCCTCCACACCTATTTATTGAGGTGCACCTGCTGGGGATTCATAATGAG AATATAACAAGAGGATCTCGGTGAAAGGCCTTAGTGGGTGTTTTGTGTGAGGTGGCTT GTAGCTAGCTACTTCCTTACAGATGGTAGAGTATTCCAATCCTCTTTGTGTTAGGGTT CTTGCTTCCAGTTTGGGATGTATTAAAACCACCATTTCACTGCTTCCCTTCCTCAATA TGCTCTGCAGCTTTTCTTGCTGTTTAAACCTCTCGCCTCAGCTTTATTTATTTGTAAG CTGCATTACTAACTGCCCAGTGATTCGGTGAAAGCTTTTTACTGAAAAAGTTAACATT TCTAGTCATCCCAATCAACTGGCTTTTTTCAACCAAAATTTTATATCATTCTTTGTCT ATCAGATACGAGAGGAAGGAAGATAATACGAAGACATGTTGAATAGTGAAAAAAAAAA AAAAGAACCACAAAAACTGGGGCAAGCCAATGTGATGTATCACTCACTGTAAGATGGC AAATGTTTTCATTTTTAAGATTCCGAATGTAAACTAGTGTGCTAGAAAGCAAACCACC CGCCACTCAAACCAGTAATTACCTTAAGCCTTAATATATTTATTAAAATACTTTATGA GAACATTACACTTTGTAGGTTAAAAATGAGGATAAAATGCTAAACTATCAAAAAAAAA AAAAAAGAAAAA
	ORF Start: ATG at 37	ORF Stop: TGA at 466
	SEQ ID NO: 122	143 aa	MW at 14441.1kD
NOV28a, CG98121-01 Protein Sequence	MSGARAAPGAAGNGAVRGLRVDGLPPLPKSLSGLLHSASGGGASGGWRHLERLYAQKS RIQDELSRGGPGGGGARAGSAARQASQPGRRSGAAPQRDGWSPPAGHVLALPTVQPLR VDSGVQGGMPGSLQPRLHLRSGERLLR		

Further analysis of the NOV28a protein yielded the following properties shown in Table 28B.

Table 28B. Protein Sequence Properties NOV28a	
PSort analysis:	0.8231 probability located in lysosome (lumen); 0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.0580 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV28a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 28C.

	Table 28C. Geneseq Results for NOV28a			
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV28a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM38970	Human polypeptide SEQ ID NO 2115 - Homo sapiens, 154 aa. [WO200153312-A1, 26-JUL-2001]	7..126 9..129	59/132 (44%) 71/132 (53%)	8e-16
AAM40756	Human polypeptide SEQ ID NO 5687 - Homo sapiens, 302 aa. [WO200153312-A1, 26-JUL-2001]	77..126 21..72	24/52 (46%) 31/52 (59%)	0.001
ABG03717	Novel human diagnostic protein #3708 - Homo sapiens, 505 aa. [WO200175067-A2, 11-OCT-2001]	2..140 15..153	45/143 (31%) 53/143 (36%)	0.019
ABG03717	Novel human diagnostic protein #3708 - Homo sapiens, 505 aa. [WO200175067-A2, 11-OCT-2001]	2..140 15..153	45/143 (31%) 53/143 (36%)	0.019
ABB11397	Human secreted protein homologue, SEQ ID NO:1767 - Homo sapiens, 686 aa. [WO200157188-A2, 09-AUG-2001]	4..136 121..253	45/148 (30%) 55/148 (36%)	0.055

5

In a BLAST search of public sequence databases, the NOV28a protein was found to have homology to the proteins shown in the BLASTP data in Table 28D.

Table 28D. Public BLASTP Results for NOV28a				
Protein Accession Number	**Protein/Organism/Length**	**NOV28a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q96GI7	UNKNOWN (PROTEIN FOR MGC:15887) - Homo sapiens (Human), 184 aa.	1..86 1..84	80/86 (93%) 80/86 (93%)	3e-38
Q9JHD8	MMTV RECEPTOR VARIANT 1 - Mus musculus (Mouse), 176 aa.	4..80 6..78	44/84 (52%) 51/84 (60%)	3e-10
Q9JL54	MMTV RECEPTOR VARIANT 2 - Mus musculus (Mouse), 189 aa.	4..80 6..78	44/84 (52%) 51/84 (60%)	3e-10
Q9QUI1	C184L ORF2 PROTEIN - Mus musculus (Mouse), 189 aa.	4..80 6..78	44/84 (52%) 51/84 (60%)	3e-10
Q9RJB0	HYPOTHETICAL 34.9 KDA PROTEIN - Streptomyces coelicolor, 324 aa.	10..107 27..103	33/98 (33%) 45/98 (45%)	0.005

PFam analysis predicts that the NOV28a protein contains the domains shown in the Table 28E.

Table 28E. Domain Analysis of NOV28a			
Pfam Domain	**NOV28a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**

5

EXAMPLE 29.

The NOV29 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 29A.

Table 29A. NOV29 Sequence Analysis		
	SEQ ID NO: 123	970 bp
NOV29a, CG99662-01 DNA Sequence	GCTGCTGGTTTTGAAACATGAATCTTTCGCTCGTCCTGGCTGCCTTTTGCTTGGGAAT AGCCTCCGCTGTTCCAAAATTTGACCAAAATTTGGATACAAAGTGGTACCAGTGGAAG GCAACACACAGAAGATTATATGGCGCGAATGAAGAAGGATGGAGGAGAGCAGTGTGGG AAAAGAATATGAAAATGATTGAACTGCACAATGGGGAATACAGCCAAGGGAAACATGG CTTCACAATGGCCATGAATGCTTTTGGTGACATGACCAATGAAGAATTCAGGCAGATG ATGGGTTGCTTTCGAAACCAGAAATTCAGGAAGGGGAAAGTGTTCCGTGAGCCTCTGT TTCTTGATCTTCCCAAATCTGTGGATTGGAGAAAGAAAGGCTACGTGACGCCAGTGAA GAATCAGAATCTGGTGGACTGTTCGCGTCCTCAAGGCAATCAGGGCTGCAATGGTGGC TTCATGGCTAGGGCCTTCCAGTATGTCAAGGAGAACGGAGGCCTGGACTCTGAGGAAT CCTATCCATATGTAGCAGTGGATGAAATCTGTAAGTACAGACCTGAGAATTCTGTTGC TAATGACACTGGCTTCACAGTGGTCGCACCTGGAAAGGAGAAGGCCCTGATGAAAGCA GTCGCAACTGTGGGGCCCATCTCCGTTGCTATGGATGCAGGCCATTCGTCCTTCCAGT TCTACAAATCAGGCATTTATTTTGAACCAGACTGCAGCAGCAAAAACCTGGATCATGG TGTTCTGGTGGTTGGCTACGGCTTTGAAGGAGCAAATTCGAATAACAGCAAGTATTGG CTCGTCAAAAACAGCTGGGGTCCAGAATGGGGCTCGAATGGCTATGTAAAAATAGCCA AAGACAAGAACAACCACTGTGGAATCGCCACAGCAGCCAGCTACCCCAATGTGTGAGC TGATGGATGGTGAGGAGGAAGGACTTAAGGACAGCATGTCTA
	ORF Start: ATG at 18	ORF Stop: TGA at 924
	SEQ ID NO: 124	302 aa	MW at 33897.2kD
NOV29a, CG99662-01 Protein Sequence	MNLSLVLAAFCLGIASAVPKFDQNLDTKWYQWKATHRRLYGANEEGWRRAVWEKNMKM IELHNGEYSQGKHGFTMAMNAFGDMTNEEFRQMMGCFRNQKFRKGKVFREPLFLDLPK SVDWRKKGYVTPVKNQNLVDCSRPQGNQGCNGGFMARAFQYVKENGGLDSEESYPYVA VDEICKYRPENSVANDTGFTVVAPGKEKALMKAVATVGPISVAMDAGHSSFQFYKSGI YFEPDCSSKNLDHGVLVVGYGFEGANSNNSKYWLVKNSWGPEWGSNGYVKIAKDKNNH CGIATAASYPNV	

Further analysis of the NOV29a protein yielded the following properties shown in Table 29B.

Table 29B. Protein Sequence Properties NOV29a	
PSort analysis:	0.8200 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1598 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 18 and 19

A search of the NOV29a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 29C.

Table 29C. Geneseq Results for NOV29a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV29a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU12177	Human PRO305 polypeptide sequence - Homo sapiens, 334 aa. [WO200140466-A2, 07-JUN-2001]	1..302 1..334	302/334 (90%) 302/334 (90%)	e-180
AAY81487	Human cathepsin L2 - Homo sapiens, 334 aa. [JP2000050886-A, 22-FEB-2000]	1..302 1..334	302/334 (90%) 302/334 (90%)	e-180
AAY02358	Polypeptide identified by the signal sequence trap method - Homo sapiens, 334 aa. [WO9918126-A1, 15-APR-1999]	1..302 1..334	302/334 (90%) 302/334 (90%)	e-180
AAW94300	Human cathepsin (LCAP) - Homo sapiens, 334 aa. [WO9900508-A1, 07-JAN-1999]	1..302 1..334	300/334 (89%) 301/334 (89%)	e-178
ABG21426	Novel human diagnostic protein #21417 - Homo sapiens, 336 aa. [WO200175067-A2, 11-OCT-2001]	1..302 2..336	298/335 (88%) 298/335 (88%)	e-175

In a BLAST search of public sequence databases, the NOV29a protein was found

to have homology to the proteins shown in the BLASTP data in Table 29D.

Table 29D. Public BLASTP Results for NOV29a				
Protein Accession Number	Protein/Organism/Length	NOV29a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O60911	Cathepsin L2 precursor (EC 3.4.22.43) (Cathepsin V) (Cathepsin U) - Homo sapiens (Human), 334 aa.	1..302 1..334	302/334 (90%) 302/334 (90%)	e-179
Q28944	Cathepsin L precursor (EC 3.4.22.15) - Sus scrofa (Pig), 334 aa.	1..302 1..334	240/335 (71%) 266/335 (78%)	e-140
P25975	Cathepsin L precursor (EC 3.4.22.15) - Bos taurus (Bovine), 334 aa.	1..302 1..334	233/335 (69%) 261/335 (77%)	e-138
P15242	Testin 1/2 precursor (CMB-22/CMB-23) - Rattus norvegicus (Rat), 333 aa.	5..302 5..333	180/330 (54%) 216/330 (64%)	1e-99
Q10991	Cathepsin L (EC 3.4.22.15) - Ovis aries (Sheep), 217 aa.	114..302 1..217	151/221 (68%) 166/221 (74%)	5e-83

PFam analysis predicts that the NOV29a protein contains the domains shown in

the Table 29E.

Table 29E. Domain Analysis of NOV29a			
Pfam Domain	NOV29a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Peptidase_C1	114..132	14/21 (67%) 19/21 (90%)	6.5e-09
Peptidase_C1	133..301	91/277 (33%) 156/277 (56%)	8.3e-90

5

Example B: Sequencing Methodology and Identification of NOVX Clones

1.   **GeneCalling™ Technology:** This is a proprietary method of performing

differential gene expression profiling between two or more samples developed at CuraGen

10   and described by Shimkets, et al., "Gene expression analysis by transcript profiling

coupled to a gene database query" Nature Biotechnology 17:198-803 (1999). cDNA was

derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents

5   that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then digested with up to as many as 120 pairs of restriction enzymes and pairs of linker-adaptors specific for each pair of restriction enzymes were ligated to the appropriate end. The restriction digestion generates a mixture of unique cDNA gene fragments. Limited PCR amplification is performed with primers

10   homologous to the linker adapter sequence where one primer is biotinylated and the other is fluorescently labeled. The doubly labeled material is isolated and the fluorescently labeled single strand is resolved by capillary gel electrophoresis. A computer algorithm compares the electropherograms from an experimental and control group for each of the restriction digestions. This and additional sequence-derived information is used to predict

15   the identity of each differentially expressed gene fragment using a variety of genetic databases. The identity of the gene fragment is confirmed by additional, gene-specific competitive PCR or by isolation and sequencing of the gene fragment.

      2.   SeqCalling™ Technology: cDNA was derived from various human samples

20   representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then sequenced using

25   CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly

30   when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

3.   **PathCalling^{TM} Technology:**

The NOVX nucleic acid sequences are derived by laboratory screening of cDNA library by the two-hybrid approach. cDNA fragments covering either the full length of the

5   DNA sequence, or part of the sequence, or both, are sequenced. In silico prediction was based on sequences available in CuraGen Corporation's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

The laboratory screening was performed using the methods summarized below:

10   cDNA libraries were derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors,

15   chemokines or steroids. The cDNA thus derived was then directionally cloned into the appropriate two-hybrid vector (Gal4-activation domain (Gal4-AD) fusion). Such cDNA libraries as well as commercially available cDNA libraries from Clontech (Palo Alto, CA) were then transferred from E.coli into a CuraGen Corporation proprietary yeast strain (disclosed in U. S. Patents 6,057,101 and 6,083,693, incorporated herein by reference in

20   their entireties).

Gal4-binding domain (Gal4-BD) fusions of a CuraGen Corportion proprietary library of human sequences was used to screen multiple Gal4-AD fusion cDNA libraries resulting in the selection of yeast hybrid diploids in each of which the Gal4-AD fusion contains an individual cDNA. Each sample was amplified using the polymerase chain

25   reaction (PCR) using non-specific primers at the cDNA insert boundaries. Such PCR product was sequenced; sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen

30   Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice

218

forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

Physical clone: the cDNA fragment derived by the screening procedure, covering the entire open reading frame is, as a recombinant DNA, cloned into pACT2 plasmid
5   (Clontech) used to make the cDNA library. The recombinant plasmid is inserted into the host and selected by the yeast hybrid diploid generated during the screening procedure by the mating of both CuraGen Corporation proprietary yeast strains N106' and YULH (U. S. Patents 6,057,101 and 6,083,693).

10   4.   RACE: Techniques based on the polymerase chain reaction such as rapid amplification of cDNA ends (RACE), were used to isolate or complete the predicted sequence of the cDNA of the invention. Usually multiple clones were sequenced from one or more human samples to derive the sequences for fragments. Various human tissue samples from different donors were used for the RACE reaction. The sequences derived
15   from these procedures were included in the SeqCalling Assembly process described in preceding paragraphs.

5.   Exon Linking: The NOVX target sequences identified in the present invention were subjected to the exon linking process to confirm the sequence. PCR primers were
20   designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such
25   primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain -
30   hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons

were gel purified, cloned and sequenced to high redundancy. The PCR product derived from exon linking was cloned into the pCR2.1 vector from Invitrogen. The resulting bacterial clone has an insert covering the entire open reading frame cloned into the pCR2.1 vector. The resulting sequences from all clones were assembled with themselves, with

5   other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported herein.

10

6.   **Physical Clone:** Exons were predicted by homology and the intron/exon boundaries were determined using standard genetic rules. Exons were further selected and refined by means of similarity determination using multiple BLAST (for example, tBlastN, BlastX, and BlastN) searches, and, in some instances, GeneScan and Grail.

15   Expressed sequences from both public and proprietary databases were also added when available to further define and complete the gene sequence. The DNA sequence was then manually corrected for apparent inconsistencies thereby obtaining the sequences encoding the full-length protein.

The PCR product derived by exon linking, covering the entire open reading frame,

20   was cloned into the pCR2.1 vector from Invitrogen to provide clones used for expression and screening purposes.

Example C: Quantitative expression analysis of clones in various cells and tissues

The quantitative expression of various clones was assessed using microtiter plates

25   containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on an Applied Biosystems ABI PRISM® 7700 or an ABI PRISM® 7900 HT Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing normal tissues and cancer cell lines), Panel 2 (containing

30   samples derived from tissues from normal and cancer sources), Panel 3 (containing cancer cell lines), Panel 4 (containing cells and cell lines from normal tissues and cells related to inflammatory conditions), Panel 5D/5I (containing human tissues and cell lines with an emphasis on metabolic diseases), AI_comprehensive_panel (containing normal tissue and

220

samples from autoimmune diseases), Panel CNSD.01 (containing central nervous system samples from normal and diseased brains) and CNS_neurodegeneration_panel (containing samples from normal and Alzheimer's diseased brains).

RNA integrity from all samples is controlled for quality by visual assessment of
5   agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

10   First, the RNA samples were normalized to reference nucleic acids such as constitutively expressed genes (for example, β-actin and GAPDH). Normalized RNA (5 ul) was converted to cDNA and analyzed by RTQ-PCR using One Step RT-PCR Master Mix Reagents (Applied Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions.

15   In other cases, non-normalized RNA samples were converted to single strand cDNA (sscDNA) using Superscript II (Invitrogen Corporation; Catalog No. 18064-147) and random hexamers according to the manufacturer's instructions. Reactions containing up to 10 μg of total RNA were performed in a volume of 20 μl and incubated for 60 minutes at 42°C. This reaction can be scaled up to 50 μg of total RNA in a final volume of
20   100 μl. sscDNA samples are then normalized to reference nucleic acids as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions.

Probes and primers were designed for each assay according to Applied Biosystems Primer Express Software package (version I for Apple Computer's Macintosh Power PC)
25   or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (Tm) range = 58°-60°C, primer optimal Tm = 59°C, maximum primer difference = 2°C, probe does not have 5'G, probe Tm must be 10°C greater than primer Tm, amplicon size 75bp to 100bp. The probes and
30   primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the

221

probe, respectively. Their final concentrations were: forward and reverse primers, 900nM each, and probe, 200nM.

PCR conditions: When working with RNA samples, normalized RNA from each tissue and each cell line was spotted in each well of either a 96 well or a 384-well PCR plate (Applied Biosystems). PCR cocktails included either a single gene specific probe and primers set, or two multiplexed probe and primers sets (a set specific for the target clone and another gene-specific set multiplexed with the target probe). PCR reactions were set up using TaqMan® One-Step RT-PCR Master Mix (Applied Biosystems, Catalog No. 4313803) following manufacturer's instructions. Reverse transcription was performed at 48°C for 30 minutes followed by amplification/PCR cycles as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100.

When working with sscDNA samples, normalized sscDNA was used as described previously for RNA samples. PCR reactions containing one or two sets of probe and primers were set up as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions. PCR amplification was performed as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were analyzed and processed as described previously.

Panels 1, 1.1, 1.2, and 1.3D

The plates for Panels 1, 1.1, 1.2 and 1.3D include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in these panels are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in these panels are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on these panels are comprised of samples derived from all major organ

systems from single adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland,

5   adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose.

In the results for Panels 1, 1.1, 1.2 and 1.3D, the following abbreviations are used:

ca. = carcinoma,
* = established from metastasis,
10   met = metastasis,
s cell var = small cell variant,
non-s = non-sm = non-small,
squam = squamous,
pl. eff = pl effusion = pleural effusion,
15   glio = glioma,
astro = astrocytoma, and
neuro = neuroblastoma.

General_screening_panel_v1.4 and General_screening_panel_v1.5

20   The plates for Panels 1.4 and 1.5 include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in Panels 1.4 and 1.5 are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate

25   cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in Panel 1.4 are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on Panels 1.4 and 1.5 are comprised of pools of samples derived from all

30   major organ systems from 2 to 5 different adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon,

35   bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose. Abbreviations are as described for Panels 1, 1.1, 1.2, and 1.3D.

Panels 2D and 2.2

223

MISSING AT THE TIME OF PUBLICATION

cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene, La Jolla, CA) and thymus and kidney (Clontech) was employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA).

5   Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial

10   cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately

15   1-5ng/ml, TNF alpha at approximately 5-10ng/ml, IFN gamma at approximately 20-50ng/ml, IL-4 at approximately 5-10ng/ml, IL-9 at approximately 5-10ng/ml, IL-13 at approximately 5-10ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen

20   Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5}M (Gibco), and 10mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20ng/ml PMA and 1-2μg/ml ionomycin, IL-12 at 5-10ng/ml, IFN gamma at

25   20-50ng/ml and IL-18 at 5-10ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5}M (Gibco), and 10mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5μg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation.

30   MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately 2×10^{6}cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco),

mercaptoethanol (5.5x10⁻⁵M) (Gibco), and 10mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions.

5   Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), and 10mM Hepes (Gibco), 50ng/ml GMCSF and 5ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids

10   (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), 10mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10μg/ml for 6 and 12-14 hours.

15   CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive

20   selection. CD45RO beads were then used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), and 10mM Hepes (Gibco) and plated at 10⁶cells/ml onto Falcon 6 well tissue culture plates

25   that had been coated overnight with 0.5μg/ml anti-CD28 (Pharmingen) and 3ug/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100μM non essential amino

30   acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), and 10mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second

expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5}M (Gibco), and 10mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

5   To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resupended at 10^6cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5}M (Gibco), and 10mM Hepes (Gibco). To activate the cells, we used PWM at 5μg/ml or anti-CD40

10   (Pharmingen) at approximately 10μg/ml and IL-4 at 5-10ng/ml. Cells were harvested for RNA preparation at 24,48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10μg/ml anti-CD28 (Pharmingen) and 2μg/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes

15   (Poietic Systems, German Town, MD) were cultured at 10^5-10^6cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5}M (Gibco), 10mM Hepes (Gibco) and IL-2 (4ng/ml). IL-12 (5ng/ml) and anti-IL4 (1μg/ml) were used to direct to Th1, while IL-4 (5ng/ml) and anti-IFN gamma (1μg/ml) were used to direct to Th2 and IL-10 at 5ng/ml was used to

20   direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5}M (Gibco), 10mM Hepes (Gibco) and IL-2 (1ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and

25   cytokines as described above, but with the addition of anti-CD95L (1μg/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third

30   activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1mM dbcAMP at

$5x10^5$cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to $5x10^5$cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol

5      $5.5x10^{-5}$M (Gibco), 10mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10ng/ml and ionomycin at 1μg/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol

10     $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5ng/ml IL-4, 5ng/ml IL-9, 5ng/ml IL-13 and 25ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately

15   10^7cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15ml Falcon Tube. An equal volume of isopropanol was added and left at -20°C overnight. The precipitated RNA was spun down

20   at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300μl of RNAse-free water and 35μl buffer (Promega) 5μl DTT, 7μl RNAsin and 8μl DNAse were added. The tube was incubated at 37°C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3M sodium acetate and 2 volumes of 100% ethanol.

25   The RNA was spun down and placed in RNAse free water. RNA was stored at -80°C.

AI_comprehensive panel_v1.0

The plates for AI_comprehensive panel_v1.0 include two control wells and 89 test samples comprised of cDNA isolated from surgical and postmortem human tissues obtained from the Backus Hospital and Clinomics (Frederick, MD). Total RNA was

30   extracted from tissue samples from the Backus Hospital in the Facility at CuraGen. Total RNA from other tissues was obtained from Clinomics.

Joint tissues including synovial fluid, synovium, bone and cartilage were obtained from patients undergoing total knee or hip replacement surgery at the Backus Hospital.

228

Tissue samples were immediately snap frozen in liquid nitrogen to ensure that isolated RNA was of optimal quality and not degraded. Additional samples of osteoarthritis and rheumatoid arthritis joint tissues were obtained from Clinomics. Normal control tissues were supplied by Clinomics and were obtained during autopsy of trauma victims.

5   Surgical specimens of psoriatic tissues and adjacent matched tissues were provided as total RNA by Clinomics. Two male and two female patients were selected between the ages of 25 and 47. None of the patients were taking prescription drugs at the time samples were isolated.

Surgical specimens of diseased colon from patients with ulcerative colitis and 10   Crohns disease and adjacent matched tissues were obtained from Clinomics. Bowel tissue from three female and three male Crohn's patients between the ages of 41-69 were used. Two patients were not on prescription medication while the others were taking dexamethasone, phenobarbital, or tylenol. Ulcerative colitis tissue was from three male and four female patients. Four of the patients were taking lebvid and two were on 15   phenobarbital.

Total RNA from post mortem lung tissue from trauma victims with no disease or with emphysema, asthma or COPD was purchased from Clinomics. Emphysema patients ranged in age from 40-70 and all were smokers, this age range was chosen to focus on patients with cigarette-linked emphysema and to avoid those patients with 20   alpha-1anti-trypsin deficiencies. Asthma patients ranged in age from 36-75, and excluded smokers to prevent those patients that could also have COPD. COPD patients ranged in age from 35-80 and included both smokers and non-smokers. Most patients were taking corticosteroids, and bronchodilators.

In the labels employed to identify tissues in the AI_comprehensive panel_v1.0 25   panel, the following abbreviations are used:

            AI = Autoimmunity
            Syn = Synovial
            Normal = No apparent disease
            Rep22 /Rep20 = individual patients
30   RA = Rheumatoid arthritis
            Backus = From Backus Hospital
            OA = Osteoarthritis
            (SS) (BA) (MF) = Individual patients
            Adj = Adjacent tissue
35   Match control = adjacent tissues
            -M = Male
            -F = Female

COPD = Chronic obstructive pulmonary disease

Panels 5D and 5I

The plates for Panel 5D and 5I include two control wells and a variety of cDNAs

5 isolated from human tissues and cell lines with an emphasis on metabolic diseases.

Metabolic tissues were obtained from patients enrolled in the Gestational Diabetes study.

Cells were obtained during different stages in the differentiation of adipocytes from

human mesenchymal stem cells. Human pancreatic islets were also obtained.

In the Gestational Diabetes study subjects are young (18 - 40 years), otherwise

10 healthy women with and without gestational diabetes undergoing routine (elective)

Caesarean section. After delivery of the infant, when the surgical incisions were being

repaired/closed, the obstetrician removed a small sample (<1 cc) of the exposed metabolic

tissues during the closure of each surgical level. The biopsy material was rinsed in sterile

saline, blotted and fast frozen within 5 minutes from the time of removal. The tissue was

15 then flash frozen in liquid nitrogen and stored, individually, in sterile screw-top tubes and

kept on dry ice for shipment to or to be picked up by CuraGen. The metabolic tissues of

interest include uterine wall (smooth muscle), visceral adipose, skeletal muscle (rectus)

and subcutaneous adipose. Patient descriptions are as follows:

Patient 2: Diabetic Hispanic, overweight, not on insulin
20 Patient 7-9: Nondiabetic Caucasian and obese (BMI>30)
Patient 10: Diabetic Hispanic, overweight, on insulin
Patient 11: Nondiabetic African American and overweight
Patient 12: Diabetic Hispanic on insulin

25 Adipocyte differentiation was induced in donor progenitor cells obtained from

Osirus (a division of Clonetics/BioWhittaker) in triplicate, except for Donor 3U which had

only two replicates. Scientists at Clonetics isolated, grew and differentiated human

mesenchymal stem cells (HuMSCs) for CuraGen based on the published protocol found in

Mark F. Pittenger, et al., Multilineage Potential of Adult Human Mesenchymal Stem Cells

30 Science Apr 2 1999: 143-147. Clonetics provided Trizol lysates or frozen pellets suitable

for mRNA isolation and ds cDNA production. A general description of each donor is as

follows:

Donor 2 and 3 U: Mesenchymal Stem cells, Undifferentiated Adipose
Donor 2 and 3 AM: Adipose, AdiposeMidway Differentiated
35 Donor 2 and 3 AD: Adipose, Adipose Differentiated

230

Human cell lines were generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: kidney proximal convoluted tubule, uterine smooth muscle cells, small intestine, liver HepG2 cancer cells, heart primary stromal cells, and adrenal cortical adenoma cells. These

5   cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. All samples were processed at CuraGen to produce single stranded cDNA.

Panel 5I contains all samples previously described with the addition of pancreatic islets from a 58 year old female patient obtained from the Diabetes Research Institute at

10   the University of Miami School of Medicine. Islet tissue was processed to total RNA at an outside source and delivered to CuraGen for addition to panel 5I.

In the labels employed to identify tissues in the 5D and 5I panels, the following abbreviations are used:

                GO Adipose = Greater Omentum Adipose
15   SK = Skeletal Muscle
                UT = Uterus
                PL = Placenta
                AD = Adipose Differentiated
                AM = Adipose Midway Differentiated
20   U = Undifferentiated Stem Cells

Panel CNSD.01

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the

25   Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains

30   from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supernuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus palladus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area

35   17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus palladus, thus this

region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

5   In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

PSP = Progressive supranuclear palsy
Sub Nigra = Substantia nigra
Glob Palladus= Globus palladus
10   Temp Pole = Temporal pole
Cing Gyr = Cingulate gyrus
BA 4 = Brodman Area 4

Panel CNS_Neurodegeneration_V1.0

15   The plates for Panel CNS_Neurodegeneration_V1.0 include two control wells and 47 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center (McLean Hospital) and the Human Brain and Spinal Fluid Resource Center (VA Greater Los Angeles Healthcare System). Brains are removed from calvaria of donors between 4 and 24 hours after death, 20   sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains six brains from Alzheimer's disease (AD) patients, and eight brains from "Normal controls" who 25   showed no evidence of dementia prior to death. The eight normal control brains are divided into two categories: Controls with no dementia and no Alzheimer's like pathology (Controls) and controls with no dementia but evidence of severe Alzheimer's like pathology, (specifically senile plaque load rated as level 3 on a scale of 0-3; 0 = no evidence of plaques, 3 = severe AD senile plaque load). Within each of these brains, the 30   following regions are represented: hippocampus, temporal cortex (Brodman Area 21), parietal cortex (Brodman area 7), and occipital cortex (Brodman area 17). These regions were chosen to encompass all levels of neurodegeneration in AD. The hippocampus is a region of early and severe neuronal loss in AD; the temporal cortex is known to show neurodegeneration in AD after the hippocampus; the parietal cortex shows moderate 35   neuronal death in the late stages of the disease; the occipital cortex is spared in AD and

232

therefore acts as a "control" region within AD patients. Not all brain regions are represented in all cases.

In the labels employed to identify tissues in the CNS_Neurodegeneration_V1.0 panel, the following abbreviations are used:

5   AD = Alzheimer's disease brain; patient was demented and showed AD-like
    pathology upon autopsy
    Control = Control brains; patient not demented, showing no neuropathology
    Control (Path) = Control brains; pateint not demented but showing sever AD-like
    pathology
10 SupTemporal Ctx = Superior Temporal Cortex
    Inf Temporal Ctx = Inferior Temporal Cortex

A. CG100041-01: Trypsin Protease

Expression of gene CG100041-01 was assessed using the primer-probe sets Ag4360 and Ag4361, described in Tables AA and AB.

15   Table AA. Probe Name Ag4360

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ggcagacactaagacgctca-3'	20	537	125
Probe	TET-5'-ggaagtttaaagacgctgggctggtt-3'-TAMRA	26	568	126
Reverse	5'-ctccactcttcctggcctag-3'	20	613	127

Table AB. Probe Name Ag4361

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-aagacgctcacagaacagga-3'	20	547	128
Probe	TET-5'-ggaagtttaaagacgctgggctggtt-3'-TAMRA	26	568	129
Reverse	5'-gtctcctccactcttcctgg-3'	20	618	130

General_screening_panel_v1.4 Summary: Ag4361 Expression of the CG100041-01 gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

20   **Panel CNS_1 Summary:** Ag4360 Expression of the CG100041-01 gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

B. CG105716-01: germline oligomeric matrix protein

Expression of gene CG105716-01 was assessed using the primer-probe set
Ag2362, described in Table BA. Results of the RTQ-PCR runs are shown in Tables BB,
BC, BD, BE, BF, BG, BH and BI.

5   <u>Table BA</u>. Probe Name Ag2362

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gtataggggatgcctgtgaca-3'	21	1205	131
Probe	TET-5'-actgtccccagaagagcaacccg-3'-TAMRA	23	1226	132
Reverse	5'-cacaagcatctcccacaaa-3'	19	1273	133

Table BB. AI_comprehensive panel_v1.0

Tissue Name	Rel. Exp.(%) Ag2362, Run 255325334	Tissue Name	Rel. Exp.(%) Ag2362, Run 255325334
110967 COPD-F	0.0	112427 Match Control Psoriasis-F	0.0
110980 COPD-F	0.0	112418 Psoriasis-M	0.0
110968 COPD-M	0.0	112723 Match Control Psoriasis-M	0.0
110977 COPD-M	0.0	112419 Psoriasis-M	0.0
110989 Emphysema-F	0.0	112424 Match Control Psoriasis-M	0.0
110992 Emphysema-F	0.3	112420 Psoriasis-M	0.0
110993 Emphysema-F	0.0	112425 Match Control Psoriasis-M	0.0
110994 Emphysema-F	0.0	104689 (MF) OA Bone-Backus	1.2
110995 Emphysema-F	0.4	104690 (MF) Adj "Normal" Bone-Backus	2.0
110996 Emphysema-F	0.1	104691 (MF) OA Synovium-Backus	2.0
110997 Asthma-M	0.0	104692 (BA) OA Cartilage-Backus	**100.0**
111001 Asthma-F	0.0	104694 (BA) OA Bone-Backus	2.0
111002 Asthma-F	0.0	104695 (BA) Adj "Normal" Bone-Backus	7.1
111003 Atopic Asthma-F	0.0	104696 (BA) OA Synovium-Backus	2.0
111004 Atopic Asthma-F	0.0	104700 (SS) OA Bone-Backus	1.8
111005 Atopic Asthma-F	0.0	104701 (SS) Adj "Normal" Bone-Backus	7.0
111006 Atopic Asthma-F	0.0	104702 (SS) OA Synovium-Backus	13.0
111417 Allergy-M	0.0	117093 OA Cartilage Rep7	0.0
112347 Allergy-M	0.0	112672 OA Bone5	0.0
112349 Normal Lung-F	0.0	112673 OA Synovium5	0.0
112357 Normal Lung-F	0.0	112674 OA Synovial Fluid cells5	0.0

112354 Normal Lung-M	0.0	117100 OA Cartilage Rep14	0.0
112374 Crohns-F	0.1	112756 OA Bone9	0.0
112389 Match Control Crohns-F	1.1	112757 OA Synovium9	0.0
112375 Crohns-F	0.0	112758 OA Synovial Fluid Cells9	0.0
112732 Match Control Crohns-F	0.0	117125 RA Cartilage Rep2	0.0
112725 Crohns-M	0.0	113492 Bone2 RA	0.0
112387 Match Control Crohns-M	0.0	113493 Synovium2 RA	0.0
112378 Crohns-M	0.0	113494 Syn Fluid Cells RA	0.0
112390 Match Control Crohns-M	0.0	113499 Cartilage4 RA	0.0
112726 Crohns-M	0.0	113500 Bone4 RA	0.0
112731 Match Control Crohns-M	0.0	113501 Synovium4 RA	0.0
112380 Ulcer Col-F	0.0	113502 Syn Fluid Cells4 RA	0.0
112734 Match Control Ulcer Col-F	0.0	113495 Cartilage3 RA	0.0
112384 Ulcer Col-F	0.1	113496 Bone3 RA	0.0
112737 Match Control Ulcer Col-F	0.0	113497 Synovium3 RA	0.0
112386 Ulcer Col-F	0.2	113498 Syn Fluid Cells3 RA	0.0
112738 Match Control Ulcer Col-F	0.0	117106 Normal Cartilage Rep20	0.0
112381 Ulcer Col-M	0.0	113663 Bone3 Normal	0.0
112735 Match Control Ulcer Col-M	0.0	113664 Synovium3 Normal	0.0
112382 Ulcer Col-M	0.2	113665 Syn Fluid Cells3 Normal	0.0
112394 Match Control Ulcer Col-M	0.0	117107 Normal Cartilage Rep22	0.0
112383 Ulcer Col-M	0.1	113667 Bone4 Normal	0.0
112736 Match Control Ulcer Col-M	0.3	113668 Synovium4 Normal	0.0
112423 Psoriasis-F	0.0	113669 Syn Fluid Cells4 Normal	0.0

Table BC. General_screening_panel_v1.5

Tissue Name	Rel. Exp.(%) Ag2362, Run 248156467	Tissue Name	Rel. Exp.(%) Ag2362, Run 248156467
Adipose	1.7	Renal ca. TK-10	0.4
Melanoma* Hs688(A).T	**100.0**	Bladder	7.5
Melanoma* Hs688(B).T	81.2	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	3.2	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	1.0	Colon ca. CaCo-2	0.1
Placenta	0.6	Colon cancer tissue	16.6
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.2
Ovarian ca. OVCAR-5	0.1	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.5
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.2
Ovary	0.5	Fetal Heart	0.1
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.7
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.2
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.1	Thymus Pool	0.1
Trachea	4.1	CNS cancer	0.0

		(glio/astro) U87-MG	
Lung	0.1	CNS cancer (glio/astro) U-118-MG	0.6
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.4	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	1.4
Lung ca. NCI-H23	0.1	Brain (fetal)	0.0
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.1
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	0.0
Fetal Liver	0.0	Brain (whole)	0.0
Liver ca. HepG2	0.8	Spinal Cord Pool	0.0
Kidney Pool	0.2	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.1
Renal ca. A498	0.0	Thyroid (female)	0.3
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	7.4

Table BD. HASS Panel v1.0

Tissue Name	Rel. Exp.(%) Ag2362, Run 268623699	Tissue Name	Rel. Exp.(%) Ag2362, Run 268623699
MCF-7 C1	0.0	U87-MG F1 (B)	0.0
MCF-7 C2	0.2	U87-MG F2	0.0
MCF-7 C3	0.4	U87-MG F3	0.0
MCF-7 C4	0.6	U87-MG F4	0.0
MCF-7 C5	0.6	U87-MG F5	0.0
MCF-7 C6	0.3	U87-MG F6	0.0
MCF-7 C7	0.0	U87-MG F7	0.0
MCF-7 C9	0.0	U87-MG F8	0.0
MCF-7 C10	0.0	U87-MG F9	0.1
MCF-7 C11	0.0	U87-MG F10	0.0
MCF-7 C12	0.0	U87-MG F11	0.0
MCF-7 C13	0.1	U87-MG F12	0.0
MCF-7 C15	0.0	U87-MG F13	0.0
MCF-7 C16	0.3	U87-MG F14	0.0
MCF-7 C17	0.0	U87-MG F15	0.0
T24 D1	0.0	U87-MG F16	0.0
T24 D2	0.2	U87-MG F17	0.4
T24 D3	0.5	LnCAP A1	0.4
T24 D4	0.0	LnCAP A2	0.0
T24 D5	0.2	LnCAP A3	0.8
T24 D6	0.0	LnCAP A4	0.5
T24 D7	0.0	LnCAP A5	0.7
T24 D9	0.0	LnCAP A6	0.3
T24 D10	0.3	LnCAP A7	2.0
T24 D11	0.0	LnCAP A8	4.9
T24 D12	0.3	LnCAP A9	2.4
T24 D13	0.3	LnCAP A10	2.6
T24 D15	0.0	LnCAP A11	6.7
T24 D16	0.9	LnCAP A12	0.3
T24 D17	0.0	LnCAP A13	0.7
CAPaN B1	0.0	LnCAP A14	1.4
CAPaN B2	0.0	LnCAP A15	3.0
CAPaN B3	0.0	LnCAP A16	0.3
CAPaN B4	0.0	LnCAP A17	5.5
CAPaN B5	0.3	Primary Astrocytes	**100.0**
CAPaN B6	0.0	Primary Renal Proximal	0.0

239

		Tubule Epithelial cell A2	
CAPaN B7	0.1	Primary melanocytes A5	0.0
CAPaN B8	0.0	126443 - 341 medullo	0.5
CAPaN B9	0.0	126444 - 487 medullo	0.0
CAPaN B10	0.0	126445 - 425 medullo	0.0
CAPaN B11	0.0	126446 - 690 medullo	1.2
CAPaN B12	0.0	126447 - 54 adult glioma	0.0
CAPaN B13	0.0	126448 - 245 adult glioma	0.3
CAPaN B14	0.3	126449 - 317 adult glioma	0.0
CAPaN B15	0.0	126450 - 212 glioma	0.0
CAPaN B16	0.0	126451 - 456 glioma	0.0
CAPaN B17	0.0		

Table BE. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2362, Run 166013008	Rel. Exp.(%) Ag2362, Run 167966893	Tissue Name	Rel. Exp.(%) Ag2362, Run 166013008	Rel. Exp.(%) Ag2362, Run 167966893
Liver adenocarcinoma	0.3	1.2	Kidney (fetal)	0.0	2.8
Pancreas	0.5	0.6	Renal ca. 786-0	0.0	0.0
Pancreatic ca. CAPAN 2	0.0	0.0	Renal ca. A498	0.0	0.1
Adrenal gland	0.3	0.0	Renal ca. RXF 393	0.0	0.0
Thyroid	0.5	1.5	Renal ca. ACHN	0.0	0.2
Salivary gland	1.8	0.2	Renal ca. UO-31	0.0	0.0
Pituitary gland	0.0	0.0	Renal ca. TK-10	0.0	0.0
Brain (fetal)	0.0	0.0	Liver	0.0	0.0
Brain (whole)	0.2	1.4	Liver (fetal)	0.0	0.0
Brain (amygdala)	0.0	0.0	Liver ca. (hepatoblast) HepG2	6.4	9.5
Brain (cerebellum)	5.4	4.6	Lung	0.9	1.6
Brain (hippocampus)	0.0	0.2	Lung (fetal)	2.2	5.3
Brain (substantia nigra)	1.6	0.9	Lung ca. (small cell) LX-1	1.6	2.0
Brain (thalamus)	0.0	0.0	Lung ca. (small cell) NCI-H69	0.0	0.0
Cerebral Cortex	0.0	0.0	Lung ca. (s.cell var.) SHP-77	0.0	0.0
Spinal cord	0.6	0.6	Lung ca. (large cell)NCI-H460	0.4	0.0
glio/astro U87-MG	0.0	0.0	Lung ca. (non-sm. cell) A549	0.0	0.3
glio/astro U-118-MG	2.3	1.6	Lung ca. (non-s.cell) NCI-H23	0.0	0.4
astrocytoma SW1783	0.0	0.0	Lung ca. (non-s.cell) HOP-62	0.0	0.0
neuro*; met SK-N-AS	0.0	0.1	Lung ca. (non-s.cl) NCI-H522	0.0	0.3
astrocytoma SF-539	0.0	0.0	Lung ca. (squam.) SW	0.0	0.0

			900		
astrocytoma SNB-75	0.0	0.3	Lung ca. (squam.) NCI-H596	0.0	0.3
glioma SNB-19	0.0	0.0	Mammary gland	2.8	1.4
glioma U251	0.0	0.0	Breast ca.* (pl.ef) MCF-7	0.0	0.0
glioma SF-295	0.3	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0	0.0
Heart (fetal)	5.4	16.5	Breast ca.* (pl.ef) T47D	0.0	0.0
Heart	4.9	12.9	Breast ca. BT-549	0.0	0.0
Skeletal muscle (fetal)	22.7	69.7	Breast ca. MDA-N	0.0	0.0
Skeletal muscle	35.4	52.5	Ovary	0.2	1.5
Bone marrow	6.4	11.2	Ovarian ca. OVCAR-3	0.0	0.0
Thymus	0.0	0.2	Ovarian ca. OVCAR-4	0.0	0.0
Spleen	0.3	0.0	Ovarian ca. OVCAR-5	0.0	0.0
Lymph node	0.3	0.0	Ovarian ca. OVCAR-8	0.0	0.0
Colorectal	0.0	0.2	Ovarian ca. IGROV-1	0.0	0.0
Stomach	1.2	0.6	Ovarian ca.* (ascites) SK-OV-3	0.0	0.0
Small intestine	0.0	0.7	Uterus	1.3	2.6
Colon ca. SW480	0.0	0.0	Placenta	20.3	1.2
Colon ca.* SW620(SW480 met)	0.3	1.5	Prostate	1.8	2.7
Colon ca. HT29	0.0	0.0	Prostate ca.* (bone met)PC-3	0.0	0.0
Colon ca. HCT-116	0.0	0.0	Testis	17.6	21.9
Colon ca. CaCo-2	0.0	0.0	Melanoma Hs688(A).T	68.8	91.4
Colon ca. tissue(ODO3866)	**100.0**	99.3	Melanoma* (met) Hs688(B).T	71.7	**100.0**
Colon ca. HCC-2998	0.0	0.0	Melanoma	0.0	0.0

			UACC-62		
Gastric ca.* (liver met) NCI-N87	0.3	0.0	Melanoma M14	0.0	0.0
Bladder	31.9	45.7	Melanoma LOX IMVI	0.0	0.0
Trachea	7.8	14.1	Melanoma* (met) SK-MEL-5	0.0	0.0
Kidney	0.3	0.7	Adipose	11.7	28.1

Table BF. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2362, Run 164151688	Tissue Name	Rel. Exp.(%) Ag2362, Run 164151688
Normal Colon	1.2	Kidney Margin 8120608	0.8
CC Well to Mod Diff (ODO3866)	54.0	Kidney Cancer 8120613	0.0
CC Margin (ODO3866)	0.9	Kidney Margin 8120614	0.9
CC Gr.2 rectosigmoid (ODO3868)	3.2	Kidney Cancer 9010320	12.8
CC Margin (ODO3868)	0.2	Kidney Margin 9010321	0.8
CC Mod Diff (ODO3920)	0.4	Normal Uterus	2.9
CC Margin (ODO3920)	0.3	Uterus Cancer 064011	0.9
CC Gr.2 ascend colon (ODO3921)	11.5	Normal Thyroid	1.7
CC Margin (ODO3921)	9.4	Thyroid Cancer 064010	1.4
CC from Partial Hepatectomy (ODO4309) Mets	46.3	Thyroid Cancer A302152	22.8
Liver Margin (ODO4309)	0.0	Thyroid Margin A302153	0.3
Colon mets to lung (OD04451-01)	2.2	Normal Breast	2.6
Lung Margin (OD04451-02)	0.1	Breast Cancer (OD04566)	21.9
Normal Prostate 6546-1	12.0	Breast Cancer (OD04590-01)	73.2
Prostate Cancer (OD04410)	36.6	Breast Cancer Mets (OD04590-03)	**100.0**
Prostate Margin (OD04410)	18.2	Breast Cancer Metastasis (OD04655-05)	9.9
Prostate Cancer (OD04720-01)	11.9	Breast Cancer 064006	35.8
Prostate Margin (OD04720-02)	2.3	Breast Cancer 1024	21.5
Normal Lung 061010	5.4	Breast Cancer 9100266	44.4

Lung Met to Muscle (ODO4286)	5.7	Breast Margin 9100265	21.3
Muscle Margin (ODO4286)	19.6	Breast Cancer A209073	44.4
Lung Malignant Cancer (OD03126)	16.0	Breast Margin A209073	4.9
Lung Margin (OD03126)	4.5	Normal Liver	0.0
Lung Cancer (OD04404)	5.2	Liver Cancer 064003	0.0
Lung Margin (OD04404)	10.7	Liver Cancer 1025	0.0
Lung Cancer (OD04565)	22.5	Liver Cancer 1026	2.9
Lung Margin (OD04565)	1.0	Liver Cancer 6004-T	0.0
Lung Cancer (OD04237-01)	15.5	Liver Tissue 6004-N	0.5
Lung Margin (OD04237-02)	12.2	Liver Cancer 6005-T	3.3
Ocular Mel Met to Liver (ODO4310)	0.3	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	0.0	Normal Bladder	52.5
Melanoma Mets to Lung (OD04321)	0.1	Bladder Cancer 1023	20.9
Lung Margin (OD04321)	0.6	Bladder Cancer A302173	19.1
Normal Kidney	0.8	Bladder Cancer (OD04718-01)	5.2
Kidney Ca, Nuclear grade 2 (OD04338)	1.1	Bladder Normal Adjacent (OD04718-03)	34.9
Kidney Margin (OD04338)	3.4	Normal Ovary	0.6
Kidney Ca Nuclear grade 1/2 (OD04339)	0.1	Ovarian Cancer 064008	71.2
Kidney Margin (OD04339)	1.3	Ovarian Cancer (OD04768-07)	0.6
Kidney Ca, Clear cell type (OD04340)	0.3	Ovary Margin (OD04768-08)	19.2
Kidney Margin (OD04340)	2.5	Normal Stomach	0.3
Kidney Ca, Nuclear grade 3 (OD04348)	2.9	Gastric Cancer 9060358	4.0
Kidney Margin (OD04348)	0.7	Stomach Margin 9060359	1.0
Kidney Cancer (OD04622-01)	3.7	Gastric Cancer 9060395	4.1
Kidney Margin (OD04622-	0.2	Stomach Margin	2.5

03)		9060394	
Kidney Cancer (OD04450-01)	0.0	Gastric Cancer 9060397	84.1
Kidney Margin (OD04450-03)	2.5	Stomach Margin 9060396	0.9
Kidney Cancer 8120607	16.8	Gastric Cancer 064005	1.1

Table BG. Panel 3D

Tissue Name	Rel. Exp.(%) Ag2362, Run 168032574	Tissue Name	Rel. Exp.(%) Ag2362, Run 168032574
Daoy- Medulloblastoma	0.0	Ca Ski- Cervical epidermoid carcinoma (metastasis)	0.0
TE671- Medulloblastoma	8.0	ES-2- Ovarian clear cell carcinoma	0.0
D283 Med- Medulloblastoma	0.6	Ramos- Stimulated with PMA/ionomycin 6h	0.0
PFSK-1- Primitive Neuroectodermal	0.0	Ramos- Stimulated with PMA/ionomycin 14h	0.6
XF-498- CNS	0.0	MEG-01- Chronic myelogenous leukemia (megokaryoblast)	0.0
SNB-78- Glioma	2.2	Raji- Burkitt's lymphoma	0.0
SF-268- Glioblastoma	0.0	Daudi- Burkitt's lymphoma	0.0
T98G- Glioblastoma	0.0	U266- B-cell plasmacytoma	0.7
SK-N-SH- Neuroblastoma (metastasis)	12.6	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	0.0	RL- non-Hodgkin's B-cell lymphoma	0.0
Cerebellum	31.6	JM1- pre-B-cell lymphoma	0.0
Cerebellum	**100.0**	Jurkat- T cell leukemia	0.0
NCI-H292- Mucoepidermoid lung carcinoma	0.0	TF-1- Erythroleukemia	0.0
DMS-114- Small cell lung cancer	57.0	HUT 78- T-cell lymphoma	0.0
DMS-79- Small cell lung cancer	29.7	U937- Histiocytic lymphoma	0.0
NCI-H146- Small cell lung cancer	6.7	KU-812- Myelogenous leukemia	0.7
NCI-H526- Small cell lung cancer	0.0	769-P- Clear cell renal carcinoma	0.0
NCI-N417- Small cell lung cancer	0.0	Caki-2- Clear cell renal carcinoma	0.0
NCI-H82- Small cell lung cancer	0.0	SW 839- Clear cell renal carcinoma	0.0

NCI-H157- Squamous cell lung cancer (metastasis)	0.0	G401- Wilms' tumor	2.3
NCI-H1155- Large cell lung cancer	0.5	Hs766T- Pancreatic carcinoma (LN metastasis)	0.6
NCI-H1299- Large cell lung cancer	0.7	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	0.0
NCI-H727- Lung carcinoid	85.9	SU86.86- Pancreatic carcinoma (liver metastasis)	1.4
NCI-UMC-11- Lung carcinoid	0.0	BxPC-3- Pancreatic adenocarcinoma	2.7
LX-1- Small cell lung cancer	11.5	HPAC- Pancreatic adenocarcinoma	0.0
Colo-205- Colon cancer	0.0	MIA PaCa-2- Pancreatic carcinoma	0.0
KM12- Colon cancer	0.0	CFPAC-1- Pancreatic ductal adenocarcinoma	0.0
KM20L2- Colon cancer	0.0	PANC-1- Pancreatic epithelioid ductal carcinoma	6.2
NCI-H716- Colon cancer	9.5	T24- Bladder carcinma (transitional cell)	0.0
SW-48- Colon adenocarcinoma	0.0	5637- Bladder carcinoma	0.0
SW1116- Colon adenocarcinoma	0.0	HT-1197- Bladder carcinoma	0.0
LS 174T- Colon adenocarcinoma	0.0	UM-UC-3- Bladder carcinma (transitional cell)	0.0
SW-948- Colon adenocarcinoma	0.0	A204- Rhabdomyosarcoma	2.4
SW-480- Colon adenocarcinoma	0.0	HT-1080- Fibrosarcoma	0.0
NCI-SNU-5- Gastric carcinoma	1.4	MG-63- Osteosarcoma	14.1
KATO III- Gastric carcinoma	0.0	SK-LMS-1- Leiomyosarcoma (vulva)	0.0
NCI-SNU-16- Gastric carcinoma	0.0	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.0
NCI-SNU-1- Gastric carcinoma	0.0	A431- Epidermoid carcinoma	0.0
RF-1- Gastric adenocarcinoma	0.0	WM266-4- Melanoma	0.5

RF-48- Gastric adenocarcinoma	4.5	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	0.0	MDA-MB-468- Breast adenocarcinoma	0.0
NCI-N87- Gastric carcinoma	0.0	SCC-4- Squamous cell carcinoma of tongue	0.0
OVCAR-5- Ovarian carcinoma	0.0	SCC-9- Squamous cell carcinoma of tongue	0.0
RL95-2- Uterine carcinoma	0.0	SCC-15- Squamous cell carcinoma of tongue	0.0
HelaS3- Cervical adenocarcinoma	1.4	CAL 27- Squamous cell carcinoma of tongue	0.0

Table BH. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2362, Run 164155977	Tissue Name	Rel. Exp.(%) Ag2362, Run 164155977
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.9
Primary Tr1 act	0.3	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.5	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.8
CD45RA CD4 lymphocyte act	4.8	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.1	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.1	Astrocytes rest	11.8
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	30.1
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.2
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	3.3

LAK cells IL-2+IL-12	0.0	Lupus kidney	26.4
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.4
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	1.7
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.8
Ramos (B cell) none	0.0	Lung fibroblast IL-9	1.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	1.2
B lymphocytes PWM	1.2	Lung fibroblast IFN gamma	0.1
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	94.6
EOL-1 dbcAMP	0.3	Dermal fibroblast CCD1070 TNF alpha	25.2
EOL-1 dbcAMP PMA/ionomycin	0.4	Dermal fibroblast CCD1070 IL-1 beta	21.6
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	76.3
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	**100.0**
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	1.8
Monocytes LPS	0.0	Colon	2.6
Macrophages rest	0.0	Lung	52.5
Macrophages LPS	0.0	Thymus	0.9
HUVEC none	0.0	Kidney	0.9
HUVEC starved	0.0		

Table BI. Panel 5D

Tissue Name	Rel. Exp.(%) Ag2362, Run 172171201	Tissue Name	Rel. Exp.(%) Ag2362, Run 172171201
97457_Patient-02go_adipose	0.3	94709_Donor 2 AM - A_adipose	17.8
97476_Patient-07sk_skeletal muscle	8.8	94710_Donor 2 AM - B_adipose	10.6
97477_Patient-07ut_uterus	0.1	94711_Donor 2 AM - C_adipose	7.9
97478_Patient-07pl_placenta	0.1	94712_Donor 2 AD - A_adipose	52.5
97481_Patient-08sk_skeletal muscle	11.0	94713_Donor 2 AD - B_adipose	73.2
97482_Patient-08ut_uterus	0.7	94714_Donor 2 AD - C_adipose	61.1
97483_Patient-08pl_placenta	0.1	94742_Donor 3 U - A_Mesenchymal Stem Cells	3.5
97486_Patient-09sk_skeletal muscle	2.2	94743_Donor 3 U - B_Mesenchymal Stem Cells	4.7
97487_Patient-09ut_uterus	0.2	94730_Donor 3 AM - A_adipose	28.5
97488_Patient-09pl_placenta	0.2	94731_Donor 3 AM - B_adipose	18.9
97492_Patient-10ut_uterus	0.6	94732_Donor 3 AM - C_adipose	19.5
97493_Patient-10pl_placenta	0.1	94733_Donor 3 AD - A_adipose	**100.0**
97495_Patient-11go_adipose	0.0	94734_Donor 3 AD - B_adipose	69.3
97496_Patient-11sk_skeletal muscle	0.1	94735_Donor 3 AD - C_adipose	82.4
97497_Patient-11ut_uterus	0.3	77138_Liver_HepG2untreated	2.6
97498_Patient-11pl_placenta	0.0	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	0.1	81735_Small Intestine	0.3
97501_Patient-12sk_skeletal muscle	0.2	72409_Kidney_Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	0.2	82685_Small intestine_Duodenum	0.0
97503_Patient-12pl_placenta	0.1	90650_Adrenal_Adrenocortical adenoma	0.0

94721_Donor 2 U - A_Mesenchymal Stem Cells	4.7	72410_Kidney_HRCE	0.0
94722_Donor 2 U - B_Mesenchymal Stem Cells	3.6	72411_Kidney_HRE	0.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	3.5	73139_Uterus_Uterine smooth muscle cells	0.0

AI_comprehensive panel_v1.0 Summary: Ag2362 Highest expression of the CG105716-01 gene is detected in cartilage from osteoarthritis patient (CT=19). In addition, high expression of this gene is also seen in synovium and bone samples from the osteoarthritis patient. Furthermore, low but significant expression of this gene is also

5   detected in synovium, bone and cartilage samples of rheumatoid arthritis patients. The CG105716-01 gene codes for cartilage oligomeric matrix protein (COMP). COMP is a noncollagenous extracellular matrix (ECM) protein which consists of five identical glycoprotein subunits, each with EGF-like and calcium-binding (thrombospondin-like) domains. COMP has been implicated in inflammatory diseases including

10   osteochondrodysplasias and arthritis (Neidhart et al., 1997, Br J Rheumatol 36(11):1151-60, PMID: 9402858; Baitner et al, 2000, J Pediatr Orthop 20(5):594-605, PMID: 11008738; Clark et al., 1999, Arthritis Rheum 1999 Nov;42(11):2356-64, PMID: 10555031). Therefore, therapeutic modulation of this gene product through the use of small molecule drugs, protein therapeutics or antibodies, might be beneficial in the

15   treatment of inflammatory diseases such as rheumatoid and osteoarthritis, and osteochondrodysplasia.

General_screening_panel_v1.5 Summary: Ag2362 Highest expression of the CG105716-01 gene is detected in melanoma sample (CT=24). Thus, expression of this gene can be used to distinguish this sample from other samples in this panel. In addition,

20   significant expression of this gene is seen in colon cancer tissue, a colon cancer, lung cancer, liver cancer, and CNS cancer cell line (CTs=31-34). The CG105716-01 gene codes for cartilage oligomeric matrix protein (COMP). Cartilage oligomeric matrix protein (COMP) is a noncollagenous extracellular matrix (ECM) protein which consists of five identical glycoprotein subunits, each with EGF-like and calcium-binding

25   (thrombospondin-like) domains. COMP contains an RGD sequence. The RGD domain in

other proteins has been shown to affect cell adhesion, migration , survival and proliferation.

Mutations of COMP can cause the osteochondrodysplasias pseudochondroplasia (PSACH) and multiple epiphyseal dysplasia (MED) (Kleerekoper et al., 2002, J Biol Chem 2002 Jan 8; [epub ahead of print], PMID: 11782471). Based on this profile, COMP may play a role in tumor cell growth and survival based upon the cells ability to interact with the extracellular matrix. Thus, therapeutic targeting with a human monoclonal antibody might block the interaction of cancer cells, or supporting stromal elements, with extracellular matrix and thus promote cell death rather than cell survival especially in these cancers. Additionally, this gene is expressed in two melanoma cell lines that mimic some of characteristics of activated tumor endothelial cells. Hence, antibody directed against this gene may affect endothelial growth and survival in the tumor and prevent tumor growth.

In addition, recently COMP has also been implicated in vascular calcification and fibrosis especially associated with with advanced complicated atherosclerosis (Canfield et al., 2002, J Pathol 196(2):228-34, PMID: 11793375). Therefore, therapeutic modulation of this gene could also be beneficial in the treatement of vascular calcification and fibrosis.

Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in pancreas, adipose, thyroid, skeletal muscle, heart, and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

HASS Panel v1.0 Summary: Ag2362 The expression of this gene appears to be highest in astrocytes (Ct=28.95). There is a slight induction in expression of this gene when LnCAP cells are serum-starved and subjected to a reduced oxygen concentration and a decreased pH. These conditions resemble those typically found in tumors and suggest that in the tumors from which LnCAp cells are derived, expression of this gene may be regulated by these conditions.

Panel 1.3D Summary: Ag2362 Two experiments with same primer and probe set are in excellent agreement, with highest expression of the CG105716-01 gene in colon cancer ODO3866 sample (CTs=29). High expression of this gene are also associated with melanoma, and a liver cancer cell line. In addition, moderate expression of this gene is

254

also seen adipose, brain, bone marrow, skeletal muscle heart, placenta, lung, testis and prostate. Please see panel 1.4 for the utility of this gene.

Panel 2D Summary: Ag2362 The expression of this gene appears to be highest in a sample derived from a breast cancer(CT=27). In addition, there appears to be substantial expression in other samples derived from breast cancer, gastric cancer, ovarian cancer, bladder cancer, thyroid cancer, kidney cancer, lung cancer, prostate cancer, liver cancer and colon cancer. Therapeutic modulation of this gene, through the use of small molecule drugs, protein therapeutics or antibodies could be of benefit in the treatment of breast, gastric, ovarian, bladder, thyroid, kidney, lung, prostate, liver or colon cancer.

Panel 3D Summary: Ag2362 Highest expression of the CG105716-01 gene is detected in cerebellum (CT=27). Low to moderate expression of this gene is associated with small cell lung cancer, lung carcinoid, and osteosarcoma. Please see panel 1.4 for the utility of this gene.

Panel 4D Summary: Ag2362 Highest expression of the CG105716-01 gene is detected in IL4 treated dermal fibroblast cells (CT=29.2). High expression of this gene is seen in all the dermal fibroblast samples (CTs=29-31). Thus expression of this gene can be used to distinguish the dermal fibroblast from other samples used in this panel. Furthermore, therapeutic modulation of this gene product could be beneficial in the treatment of skin disorders, including psoriasis.

In addition, low to moderate expression of this gene is also with lung and colon. Therefore therapeutic modulation of this gene could be useful in treatment of lung and colon related diseases such as lupus and glomerulonephritis, and inflammatory bowel diseases.

Panel 5D Summary: Ag2362 Highest expression of the CG105716-01 gene is detected in a adipose sample (CT=25). In addition, high expression of this gene is seen in other adipose samples, as well as skeletal muscle. Thus, expression of this gene could be used to distinguish this sample from other samples in this panel.

C. CG57415-01: neural cell adhesion protein

Expression of gene CG57415-01 was assessed using the primer-probe sets Ag1030, Ag3231, Ag971, Ag994 and Ag275, described in Tables CA, CB, CC, CD and CE. Results of the RTQ-PCR runs are shown in Tables CF, CG, CH, CI and CJ.

Table CA. Probe Name Ag1030

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-atggaaggcctaagcctacata-3'	22	1075	134
Probe	TET-5'-aaaatggcgaacctctgctaactcgg-3'-TAMRA	26	1108	135
Reverse	5'-ttccttgctcaatttgaattct-3'	22	1137	136

Table CB. Probe Name Ag3231

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tgctaactcgggatagaattca-3'	22	1123	137
Probe	TET-5'-tgagcaaggaacactcaacataacaa-3'-TAMRA	26	1148	138
Reverse	5'-gatacatgccagcatctgaga-3'	21	1183	139

Table CC. Probe Name Ag971

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-atggaaggcctaagcctacata-3'	22	1075	140
Probe	TET-5'-aaaatggcgaacctctgctaactcgg-3'-TAMRA	26	1108	141
Reverse	5'-ttccttgctcaatttgaattct-3'	22	1137	142

Table CD. Probe Name Ag994

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-atggaaggcctaagcctacata-3'	22	1075	143
Probe	TET-5'-aaaatggcgaacctctgctaactcgg-3'-TAMRA	26	1108	144
Reverse	5'-ttccttgctcaatttgaattct-3'	22	1137	145

Table CE. Probe Name Ag275

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ttgggaatgtaaagcaaatggaa-3'	23	1058	146
Probe	TET-5'-cctaagcctacatacaagtggctaaaaaatggcg-3'-TAMRA	34	1083	147
Reverse	5'-aattctatcccgagttagcagaggt-3'	25	1118	148

5

256

Table CF. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3231, Run 209862302	Tissue Name	Rel. Exp.(%) Ag3231, Run 209862302
AD 1 Hippo	5.6	Control (Path) 3 Temporal Ctx	4.5
AD 2 Hippo	18.9	Control (Path) 4 Temporal Ctx	35.6
AD 3 Hippo	5.6	AD 1 Occipital Ctx	13.4
AD 4 Hippo	4.9	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	97.3	AD 3 Occipital Ctx	5.3
AD 6 Hippo	18.2	AD 4 Occipital Ctx	15.9
Control 2 Hippo	21.5	AD 5 Occipital Ctx	24.5
Control 4 Hippo	3.8	AD 6 Occipital Ctx	69.7
Control (Path) 3 Hippo	3.9	Control 1 Occipital Ctx	4.3
AD 1 Temporal Ctx	10.2	Control 2 Occipital Ctx	75.3
AD 2 Temporal Ctx	33.0	Control 3 Occipital Ctx	18.6
AD 3 Temporal Ctx	6.2	Control 4 Occipital Ctx	3.1
AD 4 Temporal Ctx	18.2	Control (Path) 1 Occipital Ctx	92.0
AD 5 Inf Temporal Ctx	86.5	Control (Path) 2 Occipital Ctx	12.2
AD 5 SupTemporal Ctx	19.8	Control (Path) 3 Occipital Ctx	1.9
AD 6 Inf Temporal Ctx	39.2	Control (Path) 4 Occipital Ctx	21.6
AD 6 Sup Temporal Ctx	38.7	Control 1 Parietal Ctx	7.5
Control 1 Temporal Ctx	7.7	Control 2 Parietal Ctx	40.9
Control 2 Temporal Ctx	49.0	Control 3 Parietal Ctx	20.2
Control 3 Temporal Ctx	13.1	Control (Path) 1 Parietal Ctx	**100.0**
Control 4 Temporal	8.2	Control (Path) 2	25.5

257

Ctx		Parietal Ctx	
Control (Path) 1 Temporal Ctx	61.1	Control (Path) 3 Parietal Ctx	4.4
Control (Path) 2 Temporal Ctx	39.0	Control (Path) 4 Parietal Ctx	59.9

Table CG. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3231, Run 214440502	Tissue Name	Rel. Exp.(%) Ag3231, Run 214440502
Adipose	10.1	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	11.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.3
Melanoma* M14	2.7	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.2
Squamous cell carcinoma SCC-4	0.3	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	32.5	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	8.6	Colon ca. CaCo-2	**100.0**
Placenta	1.1	Colon cancer tissue	12.8
Uterus Pool	13.5	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	2.6	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	45.7
Ovarian ca. OVCAR-5	0.1	Small Intestine Pool	29.3
Ovarian ca. IGROV-1	0.0	Stomach Pool	25.7
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	9.7
Ovary	2.9	Fetal Heart	7.4
Breast ca. MCF-7	0.0	Heart Pool	6.8
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	29.3
Breast ca. BT 549	74.2	Fetal Skeletal Muscle	3.8
Breast ca. T47D	0.0	Skeletal Muscle Pool	2.6
Breast ca. MDA-N	0.0	Spleen Pool	2.6
Breast Pool	47.6	Thymus Pool	26.2
Trachea	11.0	CNS cancer	0.0

		(glio/astro) U87-MG	
Lung	3.1	CNS cancer (glio/astro) U-118-MG	0.1
Fetal Lung	27.9	CNS cancer (neuro;met) SK-N-AS	6.8
Lung ca. NCI-N417	9.8	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	2.4	CNS cancer (astro) SNB-75	0.1
Lung ca. NCI-H146	7.4	CNS cancer (glio) SNB-19	0.1
Lung ca. SHP-77	0.4	CNS cancer (glio) SF-295	5.6
Lung ca. A549	0.0	Brain (Amygdala) Pool	31.0
Lung ca. NCI-H526	9.1	Brain (cerebellum)	23.2
Lung ca. NCI-H23	0.1	Brain (fetal)	84.7
Lung ca. NCI-H460	0.2	Brain (Hippocampus) Pool	25.3
Lung ca. HOP-62	0.1	Cerebral Cortex Pool	49.7
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	45.1
Liver	0.3	Brain (Thalamus) Pool	40.6
Fetal Liver	10.6	Brain (whole)	34.6
Liver ca. HepG2	0.0	Spinal Cord Pool	7.9
Kidney Pool	22.2	Adrenal Gland	1.7
Fetal Kidney	35.4	Pituitary gland Pool	12.2
Renal ca. 786-0	0.0	Salivary Gland	1.0
Renal ca. A498	0.0	Thyroid (female)	23.7
Renal ca. ACHN	0.1	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	36.3

Table CH. Panel 1

Tissue Name	Rel. Exp.(%) Ag275, Run 88164405	Tissue Name	Rel. Exp.(%) Ag275, Run 88164405
Endothelial cells	0.0	Renal ca. 786-0	0.0
Endothelial cells (treated)	0.0	Renal ca. A498	1.1
Pancreas	3.2	Renal ca. RXF 393	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. ACHN	0.0
Adrenal gland	0.4	Renal ca. UO-31	0.0
Thyroid	6.6	Renal ca. TK-10	0.0
Salivary gland	0.7	Liver	2.6
Pituitary gland	1.4	Liver (fetal)	0.4
Brain (fetal)	4.4	Liver ca. (hepatoblast) HepG2	0.0
Brain (whole)	17.2	Lung	0.6
Brain (amygdala)	3.5	Lung (fetal)	1.7
Brain (cerebellum)	**100.0**	Lung ca. (small cell) LX-1	0.2
Brain (hippocampus)	3.6	Lung ca. (small cell) NCI-H69	2.0
Brain (substantia nigra)	4.9	Lung ca. (s.cell var.) SHP-77	0.0
Brain (thalamus)	16.0	Lung ca. (large cell)NCI-H460	0.2
Brain (hypothalamus)	6.4	Lung ca. (non-sm. cell) A549	0.0
Spinal cord	1.3	Lung ca. (non-s.cell) NCI-H23	0.0
glio/astro U87-MG	0.0	Lung ca. (non-s.cell) HOP-62	0.0
glio/astro U-118-MG	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SW1783	0.0	Lung ca. (squam.) SW 900	0.0
neuro*; met SK-N-AS	0.5	Lung ca. (squam.) NCI-H596	3.0
astrocytoma SF-539	0.0	Mammary gland	6.6
astrocytoma SNB-75	0.0	Breast ca.* (pl.ef) MCF-7	0.0
glioma SNB-19	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0

glioma U251	0.0	Breast ca.* (pl. ef) T47D	0.1
glioma SF-295	0.2	Breast ca. BT-549	2.8
Heart	0.9	Breast ca. MDA-N	0.0
Skeletal muscle	0.2	Ovary	1.1
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.1
Thymus	4.0	Ovarian ca. OVCAR-4	0.0
Spleen	0.2	Ovarian ca. OVCAR-5	0.0
Lymph node	1.5	Ovarian ca. OVCAR-8	0.0
Colon (ascending)	6.0	Ovarian ca. IGROV-1	0.0
Stomach	7.5	Ovarian ca. (ascites) SK-OV-3	0.0
Small intestine	4.2	Uterus	2.1
Colon ca. SW480	0.0	Placenta	3.2
Colon ca.* SW620 (SW480 met)	0.0	Prostate	1.0
Colon ca. HT29	0.0	Prostate ca.* (bone met) PC-3	0.0
Colon ca. HCT-116	0.0	Testis	64.2
Colon ca. CaCo-2	18.8	Melanoma Hs688(A).T	0.0
Colon ca. HCT-15	0.0	Melanoma* (met) Hs688(B).T	0.1
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca. * (liver met) NCI-N87	0.0	Melanoma M14	1.4
Bladder	5.4	Melanoma LOX IMVI	0.0
Trachea	2.9	Melanoma* (met) SK-MEL-5	0.0
Kidney	1.9	Melanoma SK-MEL-28	0.1
Kidney (fetal)	3.4		

Table CI. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag3231, Run 174442845	Tissue Name	Rel. Exp.(%) Ag3231, Run 174442845
Normal Colon	22.8	Kidney Margin (OD04348)	35.8
Colon cancer (OD06064)	11.9	Kidney malignant cancer (OD06204B)	2.1
Colon Margin (OD06064)	34.9	Kidney normal adjacent tissue (OD06204E)	3.8
Colon cancer (OD06159)	0.0	Kidney Cancer (OD04450-01)	**100.0**
Colon Margin (OD06159)	23.0	Kidney Margin (OD04450-03)	19.2
Colon cancer (OD06297-04)	1.2	Kidney Cancer 8120613	0.0
Colon Margin (OD06297-05)	14.1	Kidney Margin 8120614	2.5
CC Gr.2 ascend colon (ODO3921)	3.8	Kidney Cancer 9010320	0.0
CC Margin (ODO3921)	6.1	Kidney Margin 9010321	0.8
Colon cancer metastasis (OD06104)	2.5	Kidney Cancer 8120607	11.2
Lung Margin (OD06104)	5.4	Kidney Margin 8120608	0.0
Colon mets to lung (OD04451-01)	2.1	Normal Uterus	35.6
Lung Margin (OD04451-02)	18.2	Uterine Cancer 064011	19.8
Normal Prostate	5.7	Normal Thyroid	3.6
Prostate Cancer (OD04410)	9.7	Thyroid Cancer 064010	26.6
Prostate Margin (OD04410)	12.2	Thyroid Cancer A302152	12.4
Normal Ovary	0.0	Thyroid Margin A302153	24.3
Ovarian cancer (OD06283-03)	0.0	Normal Breast	19.5
Ovarian Margin (OD06283-07)	9.2	Breast Cancer (OD04566)	1.0
Ovarian Cancer 064008	15.4	Breast Cancer 1024	5.3

Ovarian cancer (OD06145)	7.3	Breast Cancer (OD04590-01)	8.4
Ovarian Margin (OD06145)	27.2	Breast Cancer Mets (OD04590-03)	26.6
Ovarian cancer (OD06455-03)	0.0	Breast Cancer Metastasis (OD04655-05)	53.2
Ovarian Margin (OD06455-07)	9.0	Breast Cancer 064006	4.5
Normal Lung	13.4	Breast Cancer 9100266	3.0
Invasive poor diff. lung adeno (ODO4945-01	0.0	Breast Margin 9100265	0.0
Lung Margin (ODO4945-03)	16.6	Breast Cancer A209073	4.3
Lung Malignant Cancer (OD03126)	0.0	Breast Margin A2090734	10.6
Lung Margin (OD03126)	6.2	Breast cancer (OD06083)	3.2
Lung Cancer (OD05014A)	1.7	Breast cancer node metastasis (OD06083)	3.3
Lung Margin (OD05014B)	9.9	Normal Liver	2.1
Lung cancer (OD06081)	0.0	Liver Cancer 1026	0.9
Lung Margin (OD06081)	9.5	Liver Cancer 1025	5.6
Lung Cancer (OD04237-01)	1.5	Liver Cancer 6004-T	4.2
Lung Margin (OD04237-02)	28.3	Liver Tissue 6004-N	1.1
Ocular Melanoma Metastasis	3.1	Liver Cancer 6005-T	0.0
Ocular Melanoma Margin (Liver)	5.9	Liver Tissue 6005-N	7.3
Melanoma Metastasis	1.7	Liver Cancer 064003	1.2
Melanoma Margin (Lung)	16.0	Normal Bladder	6.6
Normal Kidney	12.2	Bladder Cancer 1023	0.0
Kidney Ca, Nuclear grade 2 (OD04338)	33.7	Bladder Cancer A302173	1.1
Kidney Margin (OD04338)	16.5	Normal Stomach	42.3

Kidney Ca Nuclear grade 1/2 (OD04339)	26.8	Gastric Cancer 9060397	0.0
Kidney Margin (OD04339)	8.0	Stomach Margin 9060396	0.0
Kidney Ca, Clear cell type (OD04340)	4.5	Gastric Cancer 9060395	11.7
Kidney Margin (OD04340)	20.0	Stomach Margin 9060394	7.6
Kidney Ca, Nuclear grade 3 (OD04348)	5.7	Gastric Cancer 064005	5.1

Table CJ. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3231, Run 164532021	Tissue Name	Rel. Exp.(%) Ag3231, Run 164532021
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.9
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.5
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	23.3

266

LAK cells IL-2+IL-12	0.0	Lupus kidney	6.8
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	19.2	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	6.6	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.9
Monocytes rest	0.0	IBD Crohn's	6.7
Monocytes LPS	0.0	Colon	42.6
Macrophages rest	0.0	Lung	54.3
Macrophages LPS	0.0	Thymus	**100.0**
HUVEC none	0.0	Kidney	47.3
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag3231 The CG57415-01 gene is homologus to a neural cell adhesion molecule, a membrane-bound glycoprotein that plays a role in cell-cell and cell-matrix adhesion. NCAM related proteins, such as Nr-CAM, play a critical role in neurite extension. (Sakurai T. J Cell Biol 2001 Sep 17;154(6):1259-73) In

addition, NCAMs are involved in plasticity mechanisms critical for learning, memory, and regeneration and have been implicated in brain pathology, including Alzheimer's disese. (Mikkonen M. Rev Neurosci 2001;12(4):311-25) Furthermore, this gene appears to be slightly downregulated in the temporal cortex of Alzheimer's patients when compared to

5　expression in control brains. Therefore, therapeutic modulation of the expression or function of this gene may foster focal neurite outgrowth and have utility in therapeutically countering neurite degeneration of neurodegenerative diseases such as Alzheimer's, ataxias, and Parkinson's disease.

General_screening_panel_v1.4 Summary: Ag3231 The CG57415-01 gene is

10　most highly expressed in a colon cancer cell line (CT=29.4) with significant expression also seen in a breast cancer cell line. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this protein may be useful in the treatment of colon and breast

15　cancers.

Among tissues with metabolic function, this gene is expressed at moderate to low levels in pituitary, adipose, pancreas, thyroid, fetal liver and adult and fetal skeletal muscle and heart. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that disregulated

20　expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene also shows moderate to low expression in all regions of the CNS examined. Please see CNS_neurodegeneration_v1.0 for discussion of utility of this gene in the CNS.

25　**Panel 1 Summary:** Ag275 Expression of the CG57415-01 gene is highest in samples derived from cerebellum (CT = 24.5) and testis (CT = 25.1). Thus, expression of this gene may be used to distinguish cerebellum and testis from other tissues. In addition, therapeutic modulation of this gene product, either through the use of purified protein to increase levels or through antibodies or small molecule drugs to inhibit function, might be

30　of use to treat diseases of the testis, such as infertility or testicular cancer. However, expression of this gene is also detected in other samples on this panel, although expression is largely restricted to normal tissues.

In addition to the high expression seen in cerebellum, this gene is also more moderately expressed in other CNS tissues including amygdala, hippocampus, substantia nigra, thalamus, hypothalamus and spinal cord. This gene shows homology to BIG-2, an axon-associated cell adhesion molecule (AxCAM) (Yoshihara Y. J Neurobiol 28:51-69).

5   AxCAMs are critical for the development and maintenance of neural networks within the brain. In the response to injury and/or neuronal death, gene expression during the process of compensatory synaptogenesis in many ways mirrors that seen during development. Thus, the therapeutic expression of this gene or its protein product may be beneficial in the treatment of CNS injury (stroke, head trauma, spinal cord injury) or neurodegenerative

10   diseases (Alzheimer's disease, Parkinson's disease, Huntington's disease, spinocerebellar ataxia, multiple sclerosis, ALS, or any disease resulting in neuronal atrophy or death).

The 30675585_EXT3 gene is also moderately expressed in all metabolic tissues on this panel) including pancreas (CT = 29), adrenal gland (CT = 32), thyroid (CT = 28), heart (CT = 31), skeletal muscle (CT = 33), liver (CT = 30) and fetal liver (CT = 32).

15   Therefore, this gene product may have a role in cell-cell communication in these tissues and thus be an antibody target for the treatment of diseases involving any or all of these tissues.

Panel 2.2 Summary: Ag3231 Expression of the CG57415-01 gene is highest in a sample derived from a kidney cancer (CT = 32.2), although the overall levels of

20   expression are low. In addition, there is significant expression detected in samples derived from two breast cancer metastases and normal stomach. Overall this pattern of expression, suggests that this gene might be useful in distinguishing kidney, metastatic breast cancer and stomach from other tissues. In addition, therapeutic modulation of the function of this gene product might be of use in the treatment of metastatic breast cancer or kidney cancer.

25   **Panel 4D Summary:** Ag3231 The CG57415-01 gene is expressed at low levels in normal thymus, lung, kidney and colon (CTs = 31-32). Interestingly, there is lower expression in IBD colitis and Crohns disease samples as well as in lupus kidney, suggesting that this gene may play a role in these diseases. Thus, this gene may be used to distinguish normal kidney from lupus kidney as well as normal colon from colon affected

30   by IBD or Crohns disease. In addition, this gene is expressed in an untreated eosinophil (EOL) cell line; however, EOL cells treated with PMA and ionomycin express this gene at much lower levels. This gene encodes a protein that is related to BIG2, a neural adhesion molecule. Transcript expression is detected primarily in untreated tissues and is down

regulated upon inflammation. Based on t the function of BIG2 as an adhesion and signaling molecule, the 30675585_EXT3 protein may be important in the devlopment of normal organ structure and on the normal trafficking of eosinophils from the bone marrow into peripheral tissues. Therapies using the protein encoded by this transcript may

5   therefore be important in reducing inflammation or in wound healing; similar therapies using other adhesion molecules which encourge neurite outgrowth have been proposed ((Vogelezang M.G. J. Neurosci. 21: 6732-6744.) .

D. CG58504-01: ADAMTS12

Expression of gene CG58504-01 was assessed using the primer-probe set Ag2475,

10   described in Table DA. Results of the RTQ-PCR runs are shown in Tables DB, DC, DD, DE and DF.

Table DA. Probe Name Ag2475

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-agagtgacctcaatcctgttca-3'	22	1318	149
Probe	TET-5'-acgtggctgtccttctcaccagaaag-3'-TAMRA	26	1345	150
Reverse	5'-gattgaaaccagcacagatgtc-3'	22	1371	151

Table DB. HASS Panel v1.0

Tissue Name	Rel. Exp.(%) Ag2475, Run 268366853	Tissue Name	Rel. Exp.(%) Ag2475, Run 268366853
MCF-7 C1	0.1	U87-MG F1 (B)	17.3
MCF-7 C2	0.0	U87-MG F2	12.9
MCF-7 C3	0.0	U87-MG F3	17.4
MCF-7 C4	0.0	U87-MG F4	27.4
MCF-7 C5	0.0	U87-MG F5	66.0
MCF-7 C6	0.0	U87-MG F6	84.7
MCF-7 C7	0.0	U87-MG F7	9.2
MCF-7 C9	0.0	U87-MG F8	10.6
MCF-7 C10	0.0	U87-MG F9	5.4
MCF-7 C11	0.0	U87-MG F10	61.1
MCF-7 C12	0.0	U87-MG F11	87.7
MCF-7 C13	0.0	U87-MG F12	45.7
MCF-7 C15	0.0	U87-MG F13	15.7
MCF-7 C16	0.0	U87-MG F14	25.2
MCF-7 C17	0.0	U87-MG F15	16.3
T24 D1	16.2	U87-MG F16	56.6
T24 D2	14.4	U87-MG F17	73.2
T24 D3	40.1	LnCAP A1	0.0
T24 D4	28.1	LnCAP A2	0.0
T24 D5	31.4	LnCAP A3	0.0
T24 D6	23.2	LnCAP A4	0.1
T24 D7	8.0	LnCAP A5	0.0
T24 D9	7.2	LnCAP A6	0.0
T24 D10	10.3	LnCAP A7	0.0
T24 D11	9.5	LnCAP A8	0.0

T24 D12	14.8	LnCAP A9	0.1
T24 D13	3.8	LnCAP A10	0.0
T24 D15	4.6	LnCAP A11	0.0
T24 D16	6.7	LnCAP A12	0.0
T24 D17	13.0	LnCAP A13	0.1
CAPaN B1	0.0	LnCAP A14	0.0
CAPaN B2	0.0	LnCAP A15	0.0
CAPaN B3	0.1	LnCAP A16	0.0
CAPaN B4	0.2	LnCAP A17	0.1
CAPaN B5	0.1	Primary Astrocytes	**100.0**
CAPaN B6	0.2	Primary Renal Proximal Tubule Epithelial cell A2	4.4
CAPaN B7	0.0	Primary melanocytes A5	4.5
CAPaN B8	0.1	126443 - 341 medullo	0.0
CAPaN B9	0.1	126444 - 487 medullo	0.0
CAPaN B10	0.4	126445 - 425 medullo	0.3
CAPaN B11	0.1	126446 - 690 medullo	1.0
CAPaN B12	0.3	126447 - 54 adult glioma	17.6
CAPaN B13	0.1	126448 - 245 adult glioma	13.2
CAPaN B14	0.0	126449 - 317 adult glioma	0.2
CAPaN B15	0.0	126450 - 212 glioma	4.4
CAPaN B16	0.2	126451 - 456 glioma	0.3
CAPaN B17	0.3		

Table DC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2475, Run 162401130	Tissue Name	Rel. Exp.(%) Ag2475, Run 162401130
Liver adenocarcinoma	0.0	Kidney (fetal)	5.4
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.1	Renal ca. A498	8.1
Adrenal gland	0.5	Renal ca. RXF 393	5.5
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.0	Renal ca. UO-31	23.8
Pituitary gland	0.0	Renal ca. TK-10	1.5
Brain (fetal)	0.0	Liver	0.2
Brain (whole)	0.1	Liver (fetal)	0.8
Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.0	Lung	1.9
Brain (hippocampus)	0.0	Lung (fetal)	4.3
Brain (substantia nigra)	0.1	Lung ca. (small cell) LX-1	2.0
Brain (thalamus)	0.5	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.3	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	0.2	Lung ca. (large cell)NCI-H460	0.2
glio/astro U87-MG	14.9	Lung ca. (non-sm. cell) A549	0.3
glio/astro U-118-MG	2.6	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	67.8	Lung ca. (non-s.cell) HOP-62	37.1
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	2.9	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	6.5	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	1.1	Mammary gland	3.9
glioma U251	0.7	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.4	Breast ca.* (pl.ef)	17.1

		MDA-MB-231	
Heart (fetal)	7.7	Breast ca.* (pl.ef) T47D	0.0
Heart	0.5	Breast ca. BT-549	6.0
Skeletal muscle (fetal)	**100.0**	Breast ca. MDA-N	0.0
Skeletal muscle	0.3	Ovary	13.7
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	0.3	Ovarian ca. OVCAR-4	0.0
Spleen	0.2	Ovarian ca. OVCAR-5	0.0
Lymph node	0.0	Ovarian ca. OVCAR-8	3.5
Colorectal	1.3	Ovarian ca. IGROV-1	0.0
Stomach	0.3	Ovarian ca.* (ascites) SK-OV-3	0.5
Small intestine	0.4	Uterus	0.4
Colon ca. SW480	0.0	Placenta	2.0
Colon ca.* SW620(SW480 met)	0.3	Prostate	0.2
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	1.3
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	9.5
Colon ca. tissue(ODO3866)	24.8	Melanoma* (met) Hs688(B).T	22.1
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	6.7	Melanoma LOX IMVI	1.8
Trachea	0.1	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.2	Adipose	7.9

Table DD. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2475, Run 165296233	Tissue Name	Rel. Exp.(%) Ag2475, Run 165296233
Normal Colon	21.2	Kidney Margin 8120608	1.2
CC Well to Mod Diff (ODO3866)	40.3	Kidney Cancer 8120613	2.8
CC Margin (ODO3866)	4.7	Kidney Margin 8120614	4.1
CC Gr.2 rectosigmoid (ODO3868)	29.1	Kidney Cancer 9010320	42.6
CC Margin (ODO3868)	4.7	Kidney Margin 9010321	11.3
CC Mod Diff (ODO3920)	8.8	Normal Uterus	6.9
CC Margin (ODO3920)	6.6	Uterus Cancer 064011	27.5
CC Gr.2 ascend colon (ODO3921)	60.3	Normal Thyroid	2.2
CC Margin (ODO3921)	13.6	Thyroid Cancer 064010	0.8
CC from Partial Hepatectomy (ODO4309) Mets	53.2	Thyroid Cancer A302152	7.4
Liver Margin (ODO4309)	6.7	Thyroid Margin A302153	6.3
Colon mets to lung (OD04451-01)	18.3	Normal Breast	59.9
Lung Margin (OD04451-02)	5.8	Breast Cancer (OD04566)	42.0
Normal Prostate 6546-1	1.0	Breast Cancer (OD04590-01)	51.4
Prostate Cancer (OD04410)	7.4	Breast Cancer Mets (OD04590-03)	65.1
Prostate Margin (OD04410)	14.5	Breast Cancer Metastasis (OD04655-05)	4.3
Prostate Cancer (OD04720-01)	7.1	Breast Cancer 064006	67.8
Prostate Margin (OD04720-02)	14.2	Breast Cancer 1024	86.5
Normal Lung 061010	23.0	Breast Cancer 9100266	19.6

Lung Met to Muscle (ODO4286)	15.5	Breast Margin 9100265	48.3
Muscle Margin (ODO4286)	9.7	Breast Cancer A209073	93.3
Lung Malignant Cancer (OD03126)	**100.0**	Breast Margin A209073	53.6
Lung Margin (OD03126)	27.2	Normal Liver	5.8
Lung Cancer (OD04404)	78.5	Liver Cancer 064003	1.5
Lung Margin (OD04404)	25.3	Liver Cancer 1025	4.8
Lung Cancer (OD04565)	54.7	Liver Cancer 1026	16.5
Lung Margin (OD04565)	24.0	Liver Cancer 6004-T	5.7
Lung Cancer (OD04237-01)	54.7	Liver Tissue 6004-N	5.9
Lung Margin (OD04237-02)	38.7	Liver Cancer 6005-T	14.0
Ocular Mel Met to Liver (ODO4310)	0.4	Liver Tissue 6005-N	5.9
Liver Margin (ODO4310)	10.8	Normal Bladder	32.3
Melanoma Mets to Lung (OD04321)	4.7	Bladder Cancer 1023	29.7
Lung Margin (OD04321)	15.1	Bladder Cancer A302173	15.0
Normal Kidney	19.2	Bladder Cancer (OD04718-01)	48.0
Kidney Ca, Nuclear grade 2 (OD04338)	3.1	Bladder Normal Adjacent (OD04718-03)	17.6
Kidney Margin (OD04338)	6.7	Normal Ovary	9.5
Kidney Ca Nuclear grade 1/2 (OD04339)	1.4	Ovarian Cancer 064008	71.7
Kidney Margin (OD04339)	9.9	Ovarian Cancer (OD04768-07)	2.7
Kidney Ca, Clear cell type (OD04340)	4.9	Ovary Margin (OD04768-08)	10.2
Kidney Margin (OD04340)	9.9	Normal Stomach	9.0
Kidney Ca, Nuclear	17.0	Gastric Cancer	4.0

grade 3 (OD04348)		9060358	
Kidney Margin (OD04348)	6.1	Stomach Margin 9060359	2.5
Kidney Cancer (OD04622-01)	9.2	Gastric Cancer 9060395	17.2
Kidney Margin (OD04622-03)	0.8	Stomach Margin 9060394	5.7
Kidney Cancer (OD04450-01)	2.5	Gastric Cancer 9060397	56.6
Kidney Margin (OD04450-03)	9.2	Stomach Margin 9060396	1.8
Kidney Cancer 8120607	3.2	Gastric Cancer 064005	22.1

Table DE. Panel 3D

Tissue Name	Rel. Exp.(%) Ag2475, Run 164886188	Tissue Name	Rel. Exp.(%) Ag2475, Run 164886188
Daoy- Medulloblastoma	6.3	Ca Ski- Cervical epidermoid carcinoma (metastasis)	7.0
TE671- Medulloblastoma	4.9	ES-2- Ovarian clear cell carcinoma	55.1
D283 Med- Medulloblastoma	2.9	Ramos- Stimulated with PMA/ionomycin 6h	0.0
PFSK-1- Primitive Neuroectodermal	0.6	Ramos- Stimulated with PMA/ionomycin 14h	0.0
XF-498- CNS	1.3	MEG-01- Chronic myelogenous leukemia (megokaryoblast)	0.0
SNB-78- Glioma	66.0	Raji- Burkitt's lymphoma	0.0
SF-268- Glioblastoma	46.0	Daudi- Burkitt's lymphoma	0.0
T98G- Glioblastoma	18.4	U266- B-cell plasmacytoma	21.2
SK-N-SH- Neuroblastoma (metastasis)	6.0	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	0.7	RL- non-Hodgkin's B-cell lymphoma	0.0
Cerebellum	0.0	JM1- pre-B-cell lymphoma	0.0
Cerebellum	0.7	Jurkat- T cell leukemia	0.0
NCI-H292- Mucoepidermoid lung carcinoma	21.8	TF-1- Erythroleukemia	0.0
DMS-114- Small cell lung cancer	0.0	HUT 78- T-cell lymphoma	0.0
DMS-79- Small cell lung cancer	0.0	U937- Histiocytic lymphoma	0.0
NCI-H146- Small cell lung cancer	0.0	KU-812- Myelogenous leukemia	0.0
NCI-H526- Small cell lung cancer	0.0	769-P- Clear cell renal carcinoma	0.0
NCI-N417- Small cell lung cancer	0.0	Caki-2- Clear cell renal carcinoma	0.0
NCI-H82- Small cell lung cancer	0.0	SW 839- Clear cell renal carcinoma	2.8
NCI-H157- Squamous cell lung cancer (metastasis)	42.6	G401- Wilms' tumor	0.0

NCI-H1155- Large cell lung cancer	0.0	Hs766T- Pancreatic carcinoma (LN metastasis)	1.9
NCI-H1299- Large cell lung cancer	1.4	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	62.4
NCI-H727- Lung carcinoid	0.7	SU86.86- Pancreatic carcinoma (liver metastasis)	83.5
NCI-UMC-11- Lung carcinoid	0.9	BxPC-3- Pancreatic adenocarcinoma	46.3
LX-1- Small cell lung cancer	13.7	HPAC- Pancreatic adenocarcinoma	0.0
Colo-205- Colon cancer	0.0	MIA PaCa-2- Pancreatic carcinoma	0.0
KM12- Colon cancer	0.0	CFPAC-1- Pancreatic ductal adenocarcinoma	2.2
KM20L2- Colon cancer	0.0	PANC-1- Pancreatic epithelioid ductal carcinoma	0.0
NCI-H716- Colon cancer	0.0	T24- Bladder carcinma (transitional cell)	53.2
SW-48- Colon adenocarcinoma	0.0	5637- Bladder carcinoma	39.8
SW1116- Colon adenocarcinoma	0.0	HT-1197- Bladder carcinoma	1.4
LS 174T- Colon adenocarcinoma	0.0	UM-UC-3- Bladder carcinma (transitional cell)	0.0
SW-948- Colon adenocarcinoma	0.0	A204- Rhabdomyosarcoma	0.0
SW-480- Colon adenocarcinoma	0.0	HT-1080- Fibrosarcoma	0.8
NCI-SNU-5- Gastric carcinoma	0.0	MG-63- Osteosarcoma	17.3
KATO III- Gastric carcinoma	0.7	SK-LMS-1- Leiomyosarcoma (vulva)	**100.0**
NCI-SNU-16- Gastric carcinoma	3.0	SJRH30- Rhabdomyosarcoma (met to bone marrow)	47.6
NCI-SNU-1- Gastric carcinoma	0.0	A431- Epidermoid carcinoma	0.0
RF-1- Gastric adenocarcinoma	16.4	WM266-4- Melanoma	0.0
RF-48- Gastric adenocarcinoma	27.7	DU 145- Prostate carcinoma (brain metastasis)	0.0

MKN-45- Gastric carcinoma	0.6	MDA-MB-468- Breast adenocarcinoma	0.0
NCI-N87- Gastric carcinoma	0.0	SCC-4- Squamous cell carcinoma of tongue	6.1
OVCAR-5- Ovarian carcinoma	0.0	SCC-9- Squamous cell carcinoma of tongue	0.0
RL95-2- Uterine carcinoma	0.0	SCC-15- Squamous cell carcinoma of tongue	0.0
HelaS3- Cervical adenocarcinoma	0.0	CAL 27- Squamous cell carcinoma of tongue	1.0

Table DF. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2475, Run 163583185	Tissue Name	Rel. Exp.(%) Ag2475, Run 163583185
Secondary Th1 act	0.0	HUVEC IL-1beta	0.2
Secondary Th2 act	0.0	HUVEC IFN gamma	0.3
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.1
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.1
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.7
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.2
Primary Th1 rest	0.1	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	1.8
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.1
CD45RA CD4 lymphocyte act	3.9	Coronery artery SMC rest	**100.0**
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	29.9
CD8 lymphocyte act	0.0	Astrocytes rest	22.2
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	24.5
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.5
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	1.9

281

LAK cells IL-2+IL-12	0.1	Lupus kidney	0.6
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	1.7
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	1.9
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	6.2
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	1.9
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.9
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	4.3
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	7.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	20.2
Ramos (B cell) none	0.0	Lung fibroblast IL-9	13.4
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	9.2
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	7.7
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	13.4
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	51.4
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	12.9
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	3.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	37.9
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	0.2
Monocytes LPS	0.0	Colon	1.9
Macrophages rest	0.0	Lung	20.7
Macrophages LPS	0.0	Thymus	0.6
HUVEC none	0.0	Kidney	0.1
HUVEC starved	0.1		

HASS Panel v1.0 Summary: Ag2475 This gene is expressed in glioma samples and primary astrocytes in culture (highest expression CT=27.8) suggesting a role in cell growth. Expression of this gene in U87-MG (a mixed glial/astrocytoma cell line) is repressed by reducing the oxygen content of the environment. Serum starvation of these

cells induces expression. This effect is not observed in T24 (bladder cancer) cells and thus may reflect tissue specific regulation of this gene.

Panel 1.3D Summary: Ag2475 Highest expression of the CG58504-01 gene is seen in fetal skeletal muscle (CT=28.4). This expression is significantly higher than expression seen in the corresponding adult tissue (CT=36.9). Thus, expression of this gene could be used to differentiate between the fetal and adult sources of this tissue. In addition, the relative overexpression of this gene in fetal skeletal muscle suggests that the protein product may enhance muscular growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of muscle related diseases. More specifically, treatment of weak or dystrophic muscle with the protein encoded by this gene could restore muscle mass or function.

Low levels of expression are also seen in other metabolic tissues, including adipose and fetal heart, suggesting a potential role for this gene in obesity and/or diabetes.

Moderate levels of expression are also seen in cell lines derived from brain cancer, breast cancer, renal cancer, lung cancer, colon cancer and melanoma. Since cell lines and fetal tissues are, on the whole, more proliferative than normal tissues, this expression profile suggests that this gene might be involved in cell proliferation. Therefore, modulation of the expression or function of this gene may be a therapeutic avenue for the treatment of cancer or other disease that involve cell proliferation. Furthermore, therapeutic targeting of this gene product with a monoclonal antibody is anticipated to limit or block the extent of tumor cell migration and invasion and tumor metastasis, particularly in brain cancer, breast cancer, renal cancer, lung cancer, colon cancer and melanoma. This gene might also be an effective marker for the diagnosis and detection of these cancers.

Panel 2D Summary: Ag2475 Highest expression of the CG58504-01 gene is seen in a lung cancer (CT=28.3). This gene encodes a putative member of the ADAMS family. The ADAMS family of proteins has multiple domains associated with function; A fibronectin domain involved cell/extracellular matrix interaction, a thrombospondin domain involved in angiogenesis and a metalloproteinase domain involved in matrix degredation. This multi-domain structure has implications for this molecule in several tumorigenic processes, including invasion and metastasis and proliferation and cell survival. Thus, the metalloproteinase domain might play a role in cell invasion and

metastasis, the fibronectin domain may play a role in cell adhesion or survival and the thrombospondin domain might play a role in angiogenesis. ADAM 12-S cleaves insulin-like growth factor binding protein-3 (IGFBP-3). IGFBP-3 enhances the p53-dependent apoptotic response of colorectal cells to DNA damage. IGF-BP3 is inversely, associated

5   with risk for colorectal cancer. Expression of IGFBP-3 induces Growth inhibition and differentiation of the human colon carcinoma cell line, Caco-2. All these data indicate that CG58504-01 may act by cleaving and inactivating IGFBP-3 limiting its anti-tumor activity.

Thus, therapeutic targeting with a human monoclonal antibody of CG58504-01

10   may inhibit any or all of the listed activities therefore blocking the angiogenic, invasion/metastasis or growth/survival promoting activities of this molecule especially in those cancer types, like colon, lung, kidney, bladder ovarian and gastric tumors where the gene is overexpressed in the tumor compared to the normal adjacent tissue.

Panel 3D Summary: Ag2475 Highest expression of the CG58504-01 gene is seen

15   in a leiomyosarcoma cell line (CT=30.4). Significant levels of expression are also seen in other cell lines including samples derived from bladder, ovarian, lung and brain cancers. Thus, expression of this gene could be used to differentiate these samples from other samples on this panel. Please see Panel 2D for detailed discussion of utility of this gene in cancer.

20   **Panel 4D Summary:** Ag2475 Highest expression of the CG58504-01 gene is seen in resting coronary artery smooth muscle cells (CT=27.3). Moderate to low levels of expression are also seen in resting astrocytes and TNFalpha + IL-1beta treated astrocytes and coronary artery smooth muscle cells, TNF alpha and IL-4 treated dermal fibroblasts, and lung. Lower levels of expression are seen in treated and untreated lung fibroblasts.

25   This expression suggests that this gene may be a marker of smooth muscle. In addition, expression in fibroblasts and astrocytes suggests that this gene product may be involved in inflammatory conditions that involve these cells. This gene encodes a putative ADAMTS molecule which has been implicated in extracellular proteolysis and may play a critical role in the tissue degradation seen in arthritis and other inflammatory conditions. (Kuno K.

30   : J Biol Chem 1997 Jan 3;272(1):556-62) Therefore, therapeutic modulation of this gene product may be useful in the treatment of pathological and inflammatory lung and skin disorders that include chronic obstructive pulmonary disease, asthma, allergy, psoriasis and emphysema.

Panel 5 Islet Summary: Ag2475 Results from one experiment with the CG58504-01 gene are not included. The amp plot indicates that there were experimental difficulties with this run.

E. CG58586-01 and CG58586-02: CASPR4B

Expression of gene CG58586-01 and variant CG58586-02 was assessed using the primer-probe set Ag3379, described in Table EA. Results of the RTQ-PCR runs are shown in Tables EB, EC, ED and EE.

Table EA. Probe Name Ag3379

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tggaattcagcttcctttgat-3'	21	2391	152
Probe	TET-5'-ccgaggcttcatatcttcattttcct-3'-TAMRA	26	2413	153
Reverse	5'-atacatccgcgctaagttctc-3'	21	2449	154

Table EB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3379, Run 210153752	Tissue Name	Rel. Exp.(%) Ag3379, Run 210153752
AD 1 Hippo	10.4	Control (Path) 3 Temporal Ctx	5.5
AD 2 Hippo	31.2	Control (Path) 4 Temporal Ctx	35.1
AD 3 Hippo	4.7	AD 1 Occipital Ctx	17.8
AD 4 Hippo	14.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	56.3	AD 3 Occipital Ctx	4.2
AD 6 Hippo	36.3	AD 4 Occipital Ctx	28.9
Control 2 Hippo	30.1	AD 5 Occipital Ctx	27.9
Control 4 Hippo	14.2	AD 6 Occipital Ctx	19.5
Control (Path) 3 Hippo	6.1	Control 1 Occipital Ctx	4.3
AD 1 Temporal Ctx	54.0	Control 2 Occipital Ctx	37.6
AD 2 Temporal Ctx	39.0	Control 3 Occipital Ctx	15.6
AD 3 Temporal Ctx	3.9	Control 4 Occipital Ctx	19.9
AD 4 Temporal Ctx	32.5	Control (Path) 1 Occipital Ctx	79.6
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	15.5
AD 5 Sup Temporal Ctx	44.8	Control (Path) 3 Occipital Ctx	7.1
AD 6 Inf Temporal Ctx	49.7	Control (Path) 4 Occipital Ctx	14.6
AD 6 Sup Temporal Ctx	43.8	Control 1 Parietal Ctx	12.8
Control 1 Temporal Ctx	5.9	Control 2 Parietal Ctx	47.6
Control 2 Temporal Ctx	32.8	Control 3 Parietal Ctx	13.3
Control 3 Temporal Ctx	14.3	Control (Path) 1 Parietal Ctx	54.0
Control 3	16.6	Control (Path) 2	32.5

Temporal Ctx		Parietal Ctx	
Control (Path) 1 Temporal Ctx	43.8	Control (Path) 3 Parietal Ctx	5.3
Control (Path) 2 Temporal Ctx	32.5	Control (Path) 4 Parietal Ctx	48.3

Table EC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3379, Run 217043246	Tissue Name	Rel. Exp.(%) Ag3379, Run 217043246
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.1
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.3	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	1.2	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.2	Colon ca. CaCo-2	4.5
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.2	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.1
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.1
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.1
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.2
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.1
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	2.6
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.2
Breast ca. MDA-N	0.3	Spleen Pool	0.1
Breast Pool	0.0	Thymus Pool	0.2
Trachea	0.1	CNS cancer	0.0

		(glio/astro) U87-MG	
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.1	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.1
Lung ca. NCI-H146	12.1	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	11.9	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	46.7
Lung ca. NCI-H526	0.0	Brain (cerebellum)	26.6
Lung ca. NCI-H23	0.4	Brain (fetal)	74.7
Lung ca. NCI-H460	0.2	Brain (Hippocampus) Pool	49.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	49.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	54.0
Liver	0.0	Brain (Thalamus) Pool	74.2
Fetal Liver	3.3	Brain (whole)	30.6
Liver ca. HepG2	0.0	Spinal Cord Pool	**100.0**
Kidney Pool	0.0	Adrenal Gland	0.4
Fetal Kidney	0.2	Pituitary gland Pool	1.0
Renal ca. 786-0	0.0	Salivary Gland	0.1
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

Table ED. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3379, Run 165296531	Tissue Name	Rel. Exp.(%) Ag3379, Run 165296531
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.3	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.1
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.1
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	0.2

LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.1	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	42.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	**100.0**	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.1	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.4	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.3
Monocytes rest	0.0	IBD Crohn's	0.1
Monocytes LPS	0.0	Colon	0.8
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	1.7
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

Table EE. general oncology screening panel_v_2.4

Tissue Name	Rel. Exp.(%) Ag3379, Run 266889998	Tissue Name	Rel. Exp.(%) Ag3379, Run 266889998
Colon cancer 1	0.0	Bladder cancer NAT 2	0.8
Colon NAT 1	0.0	Bladder cancer NAT 3	0.0
Colon cancer 2	0.0	Bladder cancer NAT 4	0.0
Colon cancer NAT 2	0.3	Adenocarcinoma of the prostate 1	0.0
Colon cancer 3	0.9	Adenocarcinoma of the prostate 2	0.0
Colon cancer NAT 3	3.3	Adenocarcinoma of the prostate 3	0.0
Colon malignant cancer 4	1.6	Adenocarcinoma of the prostate 4	0.0
Colon normal adjacent tissue 4	1.4	Prostate cancer NAT 5	4.4
Lung cancer 1	0.0	Adenocarcinoma of the prostate 6	0.0
Lung NAT 1	0.0	Adenocarcinoma of the prostate 7	1.2
Lung cancer 2	**100.0**	Adenocarcinoma of the prostate 8	0.0
Lung NAT 2	0.4	Adenocarcinoma of the prostate 9	0.0
Squamous cell carcinoma 3	1.2	Prostate cancer NAT 10	0.0
Lung NAT 3	0.0	Kidney cancer 1	0.0
metastatic melanoma 1	0.0	KidneyNAT 1	0.0
Melanoma 2	0.0	Kidney cancer 2	2.9
Melanoma 3	0.2	Kidney NAT 2	1.0
metastatic melanoma 4	6.5	Kidney cancer 3	0.0
metastatic melanoma 5	7.6	Kidney NAT 3	1.0
Bladder cancer 1	0.0	Kidney cancer 4	0.0
Bladder cancer NAT 1	0.0	Kidney NAT 4	0.0
Bladder cancer 2	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag3379 This panel confirms the expression of the CG58586-01 gene at significant levels in the brain in an independent

292

group of individuals. This gene is found to be slighltly upregulated in the temporal cortex of Alzheimer's disease patients. Blockade of this receptor may be of use in the treatment of this disease and decrease neuronal death.

5 **General_screening_panel_v1.4 Summary:** Ag3379 Highest expression of the CG58586-01 is detected in spinal cord sample (CT=26.3). In addition, high expression of this gene is exclusively seen in all the region of central nervous system examined including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. The CG58586-01 gene codes for contactin associated protein-like 4 precursor (Cell recognition molecule, Caspr4). Caspr (paranodin) family of proteins play a

10 central role in in the assembly of multiprotein complexes necessary for the formation and maintenance of paranodal junctions (Denisenko-Nehrbass et al., 2002, J Physiol Paris 96(1-2):99-103, PMID: 11755788). Therefore, therapeutic modulation of this gene could be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

15 In addition, significant expression of this gene is seen in a colon cancer and two lung cancer cell lines. Therefore, therapeutic modulation of the activity of this gene or its protein product, through the use of small molecule drugs, protein therapeutics or antibodies, might be beneficial in the treatment of lung cancer or colon cancer.

This gene also shows moderate expression in fetal liver and skeletal muscle

20 (CTs=31). Interestingly, this gene is expressed at much higher levels in fetal when compared to adult liver and skeletal muscle (CTs=35-40). This observation suggests that expression of this gene can be used to distinguish fetal from the corresponding adult tissue. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance growth or development of liver and skeletal muscle in the

25 fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of muscle and liver related diseases.

Panel 4D Summary: Ag3379 Highest expression of the CG58586-01 is detected exclusively in Ramos B cells (CTs=27-28). Thus, expression of this gene can be used to

30 distinguish the Ramos B cells from other samples used in this panel. B cells represent a principle component of immunity and contribute to the immune response in a number of important functional roles, including antibody production. Production of antibodies against self-antigens is a major component in autoimmune disorders. In addition, low but

significant expression of this gene is also seen in thymus (CT=33). Therefore, therapeutic

modulation of this gene product may reduce or eliminate the symptoms of patients

suffering from asthma, allergies, chronic obstructive pulmonary disease, emphysema,

Crohn's disease, ulcerative colitis, rheumatoid arthritis, psoriasis, osteoarthritis, systemic

5   lupus erythematosus and other autoimmune disorders.

In addition, low but significant expression of this gene is also seen in colon

samples (CT=34). Interestingly, expression of this gene is decreased in colon samples

from patients with IBD colitis and Crohn's disease (CTs=36-37)relative to normal colon.

Therefore, therapeutic modulation of the activity of the protein encoded by this gene may

10   be useful in the treatment of inflammatory bowel disease.

general oncology screening panel_v_2.4 Summary: Ag3379 Highest expression

of the CG58586-01 is detected exclusively in lung cancer (OD06850-03C) sample

(CT=30.5). In addition, low levels of expression of this gene is also seen in two of the

metastatic melanoma samples. Therefore, expression of this gene may be used as

15   diagnostic marker for detection of lung and metastatic melanoma. Furthermore,

therapeutic modulation of this gene product may be beneficial in the treatment of lung

cancer and melanoma.

F. CG93453-01 and CG93453-02: ADAM-TS 3 PRECURSOR (KIAA0366)

20   Expression of gene CG93453-01 and full length clone CG93453-02 was assessed

using the primer-probe set Ag2085, described in Table FA. Results of the RTQ-PCR runs

are shown in Tables FB, FC and FD.

Table FA. Probe Name Ag2085

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ctgtggaagttctggctatcag-3'	22	2848	155
Probe	TET-5'-actgtacgctgccttcagccactcct-3'-TAMRA	26	2876	156
Reverse	5'-gtcacccatgcagtatttgc-3'	20	2928	157

25

Table FB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2085, Run 165627525	Tissue Name	Rel. Exp.(%) Ag2085, Run 165627525
Liver adenocarcinoma	0.9	Kidney (fetal)	11.6
Pancreas	0.0	Renal ca. 786-0	13.3
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	4.4
Adrenal gland	2.2	Renal ca. RXF 393	8.7
Thyroid	3.6	Renal ca. ACHN	**100.0**
Salivary gland	4.0	Renal ca. UO-31	46.7
Pituitary gland	0.0	Renal ca. TK-10	9.9
Brain (fetal)	22.1	Liver	0.0
Brain (whole)	11.0	Liver (fetal)	5.8
Brain (amygdala)	9.8	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.0	Lung	3.3
Brain (hippocampus)	13.4	Lung (fetal)	4.2
Brain (substantia nigra)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	2.7	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	15.2	Lung ca. (s.cell var.) SHP-77	1.5
Spinal cord	0.0	Lung ca. (large cell)NCI-H460	0.1
glio/astro U87-MG	2.3	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	4.1	Lung ca. (non-s.cell) NCI-H23	14.5
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	8.5
neuro*; met SK-N-AS	17.7	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	5.8	Lung ca. (squam.) SW 900	5.5
astrocytoma SNB-75	6.9	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	9.8	Mammary gland	8.6
glioma U251	17.0	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	2.4	Breast ca.* (pl.ef)	5.0

		MDA-MB-231	
Heart (fetal)	1.3	Breast ca.* (pl.ef) T47D	0.0
Heart	1.4	Breast ca. BT-549	12.8
Skeletal muscle (fetal)	1.3	Breast ca. MDA-N	9.9
Skeletal muscle	0.0	Ovary	1.1
Bone marrow	1.3	Ovarian ca. OVCAR-3	15.9
Thymus	3.9	Ovarian ca. OVCAR-4	0.0
Spleen	3.9	Ovarian ca. OVCAR-5	0.0
Lymph node	1.3	Ovarian ca. OVCAR-8	2.6
Colorectal	0.0	Ovarian ca. IGROV-1	0.0
Stomach	3.9	Ovarian ca.* (ascites) SK-OV-3	1.2
Small intestine	3.1	Uterus	9.2
Colon ca. SW480	0.0	Placenta	3.2
Colon ca.* SW620(SW480 met)	0.0	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.9
Colon ca. HCT-116	0.0	Testis	22.1
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	1.8
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	16.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	2.1	Melanoma* (met) SK-MEL-5	0.0
Kidney	4.0	Adipose	6.2

Table FC. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2085, Run 152862061	Tissue Name	Rel. Exp.(%) Ag2085, Run 152862061
Normal Colon	34.6	Kidney Margin 8120608	4.4
CC Well to Mod Diff (ODO3866)	10.3	Kidney Cancer 8120613	0.0
CC Margin (ODO3866)	8.7	Kidney Margin 8120614	1.6
CC Gr.2 rectosigmoid (ODO3868)	2.6	Kidney Cancer 9010320	31.9
CC Margin (ODO3868)	1.2	Kidney Margin 9010321	6.6
CC Mod Diff (ODO3920)	1.0	Normal Uterus	26.6
CC Margin (ODO3920)	12.5	Uterus Cancer 064011	33.9
CC Gr.2 ascend colon (ODO3921)	6.7	Normal Thyroid	16.6
CC Margin (ODO3921)	3.8	Thyroid Cancer 064010	26.8
CC from Partial Hepatectomy (ODO4309) Mets	4.4	Thyroid Cancer A302152	13.5
Liver Margin (ODO4309)	6.6	Thyroid Margin A302153	48.6
Colon mets to lung (OD04451-01)	1.7	Normal Breast	35.6
Lung Margin (OD04451-02)	6.2	Breast Cancer (OD04566)	1.7
Normal Prostate 6546-1	6.2	Breast Cancer (OD04590-01)	15.4
Prostate Cancer (OD04410)	21.2	Breast Cancer Mets (OD04590-03)	17.7
Prostate Margin (OD04410)	35.6	Breast Cancer Metastasis (OD04655-05)	12.1
Prostate Cancer (OD04720-01)	11.0	Breast Cancer 064006	8.9
Prostate Margin (OD04720-02)	29.9	Breast Cancer 1024	14.9
Normal Lung 061010	17.1	Breast Cancer 9100266	12.9

Lung Met to Muscle (ODO4286)	6.9	Breast Margin 9100265	5.8
Muscle Margin (ODO4286)	4.8	Breast Cancer A209073	33.4
Lung Malignant Cancer (OD03126)	27.5	Breast Margin A209073	16.0
Lung Margin (OD03126)	18.6	Normal Liver	6.5
Lung Cancer (OD04404)	28.1	Liver Cancer 064003	1.8
Lung Margin (OD04404)	17.9	Liver Cancer 1025	2.5
Lung Cancer (OD04565)	2.2	Liver Cancer 1026	7.0
Lung Margin (OD04565)	7.6	Liver Cancer 6004-T	3.7
Lung Cancer (OD04237-01)	20.4	Liver Tissue 6004-N	0.6
Lung Margin (OD04237-02)	26.1	Liver Cancer 6005-T	2.2
Ocular Mel Met to Liver (ODO4310)	4.1	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	2.1	Normal Bladder	30.6
Melanoma Mets to Lung (OD04321)	3.3	Bladder Cancer 1023	0.9
Lung Margin (OD04321)	24.1	Bladder Cancer A302173	31.4
Normal Kidney	21.9	Bladder Cancer (OD04718-01)	7.9
Kidney Ca, Nuclear grade 2 (OD04338)	62.0	Bladder Normal Adjacent (OD04718-03)	24.8
Kidney Margin (OD04338)	17.2	Normal Ovary	4.2
Kidney Ca Nuclear grade 1/2 (OD04339)	**100.0**	Ovarian Cancer 064008	27.9
Kidney Margin (OD04339)	11.4	Ovarian Cancer (OD04768-07)	13.6
Kidney Ca, Clear cell type (OD04340)	3.6	Ovary Margin (OD04768-08)	24.0
Kidney Margin (OD04340)	25.0	Normal Stomach	8.0
Kidney Ca, Nuclear	11.7	Gastric Cancer	5.0

grade 3 (OD04348)		9060358	
Kidney Margin (OD04348)	19.6	Stomach Margin 9060359	5.4
Kidney Cancer (OD04622-01)	7.6	Gastric Cancer 9060395	20.7
Kidney Margin (OD04622-03)	2.0	Stomach Margin 9060394	2.8
Kidney Cancer (OD04450-01)	17.4	Gastric Cancer 9060397	9.0
Kidney Margin (OD04450-03)	28.9	Stomach Margin 9060396	1.2
Kidney Cancer 8120607	13.3	Gastric Cancer 064005	8.0

Table FD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2085, Run 161905847	Tissue Name	Rel. Exp.(%) Ag2085, Run 161905847
Secondary Th1 act	0.0	HUVEC IL-1beta	0.3
Secondary Th2 act	0.0	HUVEC IFN gamma	0.7
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	1.3
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	1.4
Secondary Th2 rest	0.0	HUVEC IL-11	0.2
Secondary Tr1 rest	0.0	Lung Microvascular EC none	2.1
Primary Th1 act	0.6	Lung Microvascular EC TNFalpha + IL-1beta	1.3
Primary Th2 act	0.1	Microvascular Dermal EC none	0.2
Primary Tr1 act	0.3	Microsvasular Dermal EC TNFalpha + IL-1beta	0.6
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.7
Primary Th2 rest	0.1	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.2
CD45RA CD4 lymphocyte act	0.3	Coronery artery SMC rest	0.6
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.5
CD8 lymphocyte act	0.0	Astrocytes rest	0.7
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	3.1
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	11.9
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	**100.0**
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.5
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.2
LAK cells IL-2	0.0	Liver cirrhosis	0.7

300

LAK cells IL-2+IL-12	0.0	Lupus kidney	0.3
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.1
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.7
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.5
PBMC rest	0.0	Lung fibroblast none	0.6
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.9
PBMC PHA-L	0.0	Lung fibroblast IL-4	3.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	1.2
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	1.6
B lymphocytes PWM	1.3	Lung fibroblast IFN gamma	1.3
B lymphocytes CD40L and IL-4	0.1	Dermal fibroblast CCD1070 rest	1.7
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	1.9
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	1.6
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	11.6
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	21.5
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.1
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	0.4
Macrophages rest	0.0	Lung	1.9
Macrophages LPS	0.0	Thymus	0.8
HUVEC none	0.6	Kidney	1.3
HUVEC starved	0.7		

Panel 1.3D Summary: Ag2085 Highest expression of the CG93453-01 gene, an ADAM TS3 homolog, is seen in a renal cancer cell line (CT=30.1). Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel. Low but significant levels of expression are also seen in cell lines derived from

brain, lung, breast, ovarian, and melanoma cancers. Thus, therapeutic modulation of the expression or function of this gene may also be effective in the treatment of these cancers.

This gene is also expressed at low levels in the CNS, including the hippocampus, and amygdala. Therefore, therapeutic modulation of the expression or function of this

5   gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

Among tissues with metabolic function, this gene is expressed at low levels in adipose, thyroid, and fetal liver. This expression suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that disregulated expression

10   of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

Panel 2D Summary: Ag2085 Highest expression of the CG93453-01 gene is seen in a kidney cancer (CT=29.5), in agreement with expression in Panel 1.3D. Significant levels of expression are also seen in kidney cancer, breast cancer and gastric cancer. Thus,

15   expression of this gene could be used to differentiate between the renal cancer and other samples on this panel, especially normal kidney tissue. The ADAMS family of proteins has multiple domains associated with function, including a thrombospondin domain involved in angiogenesis and a metalloproteinase domain involved in matrix degredation. This multi-domain structure has implications for this molecule in several tumorigenic

20   processes, including invasion and metastasis and proliferation and cell survival. Thus, the metalloproteinase domain might play a role in cell invasion and metastasis, and the thrombospondin domain might play a role in angiogenesis. Therefore, therapeutic modulation of the expression or function of this gene may also be effective in the treatment of kidney cancer, breast cancer and gastric cancer.

25   **Panel 4D Summary:** Ag2085 Highest expression of the CG93453-01 gene is seen in the KU-812 basophil cell line treated with PMA/ionomycin (CT=26.3). This transcript appears to be induced in the PMA and ionomycin treated basophil cell line, when compared to expression in resting basophils (CT=29.4). Basophils release histamines and other biological modifiers in reponse to allergens and play an important role in the

30   pathology of asthma and hypersensitivity reactions. In addition, this gene encodes a putative ADAMTS molecule which has been implicated in extracellular proteolysis and may play a critical role in the tissue degradation seen in arthritis and other inflammatory conditions. (Kuno K. : J Biol Chem 1997 Jan 3;272(1):556-62) Therefore, therapeutics

designed against the putative protein encoded by this gene may reduce or inhibit inflammation by blocking basophil function in these diseases. In addition, these cells are a reasonable model for the inflammatory cells that take part in various inflammatory lung and bowel diseases, such as asthma, Crohn's disease, and ulcerative colitis. Therefore,

5   therapeutics that modulate the function of this gene product may reduce or eliminate the symptoms of patients suffering from asthma, Crohn's disease, and ulcerative colitis.

G. CG95145-01: C1q-related Gliacolin

Expression of gene CG95145-01 was assessed using the primer-probe set Ag4503, described in Table GA. Results of the RTQ-PCR runs are shown in Tables GB, GC and

10   GD.

Table GA. Probe Name Ag4503

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-aacctcggcaatcactatgac-3'	21	597	158
Probe	TET-5'-ctgccaggtacgcggcatctactt-3'-TAMRA	24	638	159
Reverse	5'-catgaggatgtggtaggtgaag-3'	22	662	160

Table GB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag4503, Run 224704539	Rel. Exp.(%) Ag4503, Run 230510315	Tissue Name	Rel. Exp.(%) Ag4503, Run 224704539	Rel. Exp.(%) Ag4503, Run 230510315
AD 1 Hippo	16.0	16.7	Control (Path) 3 Temporal Ctx	11.6	10.1
AD 2 Hippo	31.2	33.9	Control (Path) 4 Temporal Ctx	42.9	35.1
AD 3 Hippo	19.3	16.5	AD 1 Occipital Ctx	6.4	4.2
AD 4 Hippo	7.3	4.3	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 hippo	34.2	31.0	AD 3 Occipital Ctx	5.4	6.2
AD 6 Hippo	40.3	34.6	AD 4 Occipital Ctx	12.2	13.3
Control 2 Hippo	47.0	47.6	AD 5 Occipital Ctx	10.3	31.4
Control 4 Hippo	9.7	8.3	AD 6 Occipital Ctx	42.0	0.5
Control (Path) 3 Hippo	12.7	9.5	Control 1 Occipital Ctx	5.3	4.1
AD 1 Temporal Ctx	9.4	9.7	Control 2 Occipital Ctx	25.9	26.6
AD 2 Temporal Ctx	45.7	31.4	Control 3 Occipital Ctx	14.2	15.8
AD 3 Temporal Ctx	9.1	8.5	Control 4 Occipital Ctx	5.3	5.3
AD 4 Temporal Ctx	22.2	20.9	Control (Path) 1	58.6	45.7

			Occipital Ctx		
AD 5 Inf Temporal Ctx	59.0	50.0	Control (Path) 2 Occipital Ctx	11.7	8.2
AD 5 SupTemporal Ctx	68.3	69.3	Control (Path) 3 Occipital Ctx	1.4	1.5
AD 6 Inf Temporal Ctx	29.3	18.2	Control (Path) 4 Occipital Ctx	9.6	12.9
AD 6 Sup Temporal Ctx	29.3	28.3	Control 1 Parietal Ctx	10.7	9.4
Control 1 Temporal Ctx	16.7	15.4	Control 2 Parietal Ctx	30.4	29.9
Control 2 Temporal Ctx	39.2	35.4	Control 3 Parietal Ctx	12.5	8.6
Control 3 Temporal Ctx	21.2	19.9	Control (Path) 1 Parietal Ctx	**100.0**	**100.0**
Control 4 Temporal Ctx	16.0	13.1	Control (Path) 2 Parietal Ctx	24.1	22.2
Control (Path) 1 Temporal Ctx	84.7	78.5	Control (Path) 3 Parietal Ctx	5.3	3.4
Control (Path) 2 Temporal Ctx	48.0	42.6	Control (Path) 4 Parietal Ctx	27.9	36.1

Table GC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag4503, Run 222695711	Tissue Name	Rel. Exp.(%) Ag4503, Run 222695711
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	2.5
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.3
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.4	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.5	Colon ca. CaCo-2	0.2
Placenta	0.3	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	1.5
Ovarian ca. OVCAR-5	0.1	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	8.8	Fetal Heart	0.5
Breast ca. MCF-7	0.0	Heart Pool	0.2
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.1
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.5	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.9	Thymus Pool	0.2
Trachea	0.3	CNS cancer	0.2

		(glio/astro) U87-MG	
Lung	0.5	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.3	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.4	CNS cancer (glio) SNB-19	0.4
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.5
Lung ca. A549	0.0	Brain (Amygdala) Pool	29.7
Lung ca. NCI-H526	17.3	Brain (cerebellum)	28.5
Lung ca. NCI-H23	2.5	Brain (fetal)	62.9
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	68.8
Lung ca. HOP-62	0.1	Cerebral Cortex Pool	97.3
Lung ca. NCI-H522	0.6	Brain (Substantia nigra) Pool	51.1
Liver	0.0	Brain (Thalamus) Pool	67.4
Fetal Liver	0.0	Brain (whole)	**100.0**
Liver ca. HepG2	0.0	Spinal Cord Pool	11.3
Kidney Pool	0.3	Adrenal Gland	0.4
Fetal Kidney	0.0	Pituitary gland Pool	0.3
Renal ca. 786-0	0.0	Salivary Gland	0.2
Renal ca. A498	0.3	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

Table GD. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4503, Run 197089620	Tissue Name	Rel. Exp.(%) Ag4503, Run 197089620
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	1.1
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	2.1
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	**100.0**
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	27.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	2.5	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	5.1
LAK cells IL-2	0.0	Liver cirrhosis	0.0

LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	5.4
LAK cells PMA/ionomycin	1.5	NCI-H292 IL-13	2.6
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	3.2	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	5.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	2.3	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	12.7	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	2.1	Thymus	0.0
HUVEC none	5.1	Kidney	9.2
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag4503 Two experiments with the same probe and primer set produce results that are in excellent agreement. Highest expression of the CG54503-05 gene is seen in the parietal cortex of a control patient (CTs=27-28.6). This protein is found to be down-regulated in the temporal cortex of

5   Alzheimer's disease patients. This protein appears to be a member of the complement

family, specifically containing a C1q domain and homology to C1q-related factor. C1q is a subunit of the complex that activates the complement system. The complement system has been implicated in Alzheimer's disease because complement proteins are found in senile plaques and neuroinflammation in response to plaques appears to be a major cause of neuronal death in AD. Therefore, up-regulation of this gene or its protein product may be of use in reversing the dementia/memory loss and neuronal death associated with this disease.

References:

Lue LF, Rydel R, Brigham EF, Yang LB, Hampel H, Murphy GM Jr, Brachova L, Yan SD, Walker DG, Shen Y, Rogers J. Inflammatory repertoire of Alzheimer's disease and nondemented elderly microglia in vitro. Glia 2001 Jul;35(1):72-9

H. CG95250-01 and CG95250-02: Aminopeptidase N - isoform 2

Expression of gene CG95250-01 and variant CG95250-02 was assessed using the primer-probe sets Ag1355 and Ag4501, described in Tables HA and HB. Results of the RTQ-PCR runs are shown in Tables HC, HD, HE, HF, HG and HH. Please note that the probe and primer set Ag4501 corresponds to the CG95250-02 variant only.

Table HA. Probe Name Ag1355

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tggattgtccctattctttgg-3'	21	1901	161
Probe	TET-5'-cacaacctttagtctggctagatcaaagca-3'-TAMRA	30	1938	162
Reverse	5'-gcatttctgggaatactttgc-3'	21	1968	163

Table HB. Probe Name Ag4501

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ttcaacttgcttatgcaatgag-3'	22	2505	164
Probe	TET-5'-tgcagcaaagacccatggatacttaaca-3'-TAMRA	28	2528	165
Reverse	5'-gtgctgatggcatactccatat-3'	22	2559	166

Table HC. AI_comprehensive panel_v1.0

Tissue Name	Rel. Exp.(%) Ag1355, Run 248105758	Tissue Name	Rel. Exp.(%) Ag1355, Run 248105758
110967 COPD-F	16.6	112427 Match Control Psoriasis-F	65.5
110980 COPD-F	35.6	112418 Psoriasis-M	19.9
110968 COPD-M	16.8	112723 Match Control Psoriasis-M	8.6
110977 COPD-M	42.0	112419 Psoriasis-M	34.9
110989 Emphysema-F	66.0	112424 Match Control Psoriasis-M	12.6
110992 Emphysema-F	31.2	112420 Psoriasis-M	**100.0**
110993 Emphysema-F	16.7	112425 Match Control Psoriasis-M	80.7
110994 Emphysema-F	8.5	104689 (MF) OA Bone-Backus	77.9
110995 Emphysema-F	42.9	104690 (MF) Adj "Normal" Bone-Backus	56.3
110996 Emphysema-F	8.3	104691 (MF) OA Synovium-Backus	34.9
110997 Asthma-M	6.6	104692 (BA) OA Cartilage-Backus	0.2
111001 Asthma-F	30.8	104694 (BA) OA Bone-Backus	46.0
111002 Asthma-F	28.7	104695 (BA) Adj "Normal" Bone-Backus	48.3
111003 Atopic Asthma-F	37.1	104696 (BA) OA Synovium-Backus	66.9
111004 Atopic Asthma-F	40.6	104700 (SS) OA Bone-Backus	56.3
111005 Atopic Asthma-F	24.5	104701 (SS) Adj "Normal" Bone-Backus	52.1
111006 Atopic Asthma-F	6.0	104702 (SS) OA Synovium-Backus	65.5
111417 Allergy-M	28.1	117093 OA Cartilage Rep7	64.2
112347 Allergy-M	0.6	112672 OA Bone5	71.2
112349 Normal Lung-F	0.2	112673 OA Synovium5	24.8

112357 Normal Lung-F	31.2	112674 OA Synovial Fluid cells5	27.5
112354 Normal Lung-M	12.4	117100 OA Cartilage Rep14	8.4
112374 Crohns-F	12.5	112756 OA Bone9	3.0
112389 Match Control Crohns-F	10.1	112757 OA Synovium9	2.0
112375 Crohns-F	16.0	112758 OA Synovial Fluid Cells9	19.3
112732 Match Control Crohns-F	1.3	117125 RA Cartilage Rep2	28.3
112725 Crohns-M	8.8	113492 Bone2 RA	7.3
112387 Match Control Crohns-M	20.0	113493 Synovium2 RA	2.4
112378 Crohns-M	1.0	113494 Syn Fluid Cells RA	4.7
112390 Match Control Crohns-M	73.2	113499 Cartilage4 RA	5.4
112726 Crohns-M	29.1	113500 Bone4 RA	7.5
112731 Match Control Crohns-M	30.4	113501 Synovium4 RA	2.7
112380 Ulcer Col-F	42.0	113502 Syn Fluid Cells4 RA	3.1
112734 Match Control Ulcer Col-F	5.7	113495 Cartilage3 RA	3.0
112384 Ulcer Col-F	85.3	113496 Bone3 RA	5.3
112737 Match Control Ulcer Col-F	11.0	113497 Synovium3 RA	2.4
112386 Ulcer Col-F	24.5	113498 Syn Fluid Cells3 RA	5.3
112738 Match Control Ulcer Col-F	2.4	117106 Normal Cartilage Rep20	3.8
112381 Ulcer Col-M	3.3	113663 Bone3 Normal	5.3
112735 Match Control Ulcer Col-M	60.7	113664 Synovium3 Normal	0.3
112382 Ulcer Col-M	15.6	113665 Syn Fluid Cells3 Normal	2.5
112394 Match Control Ulcer Col-M	14.1	117107 Normal Cartilage Rep22	22.7
112383 Ulcer Col-M	39.5	113667 Bone4 Normal	27.7
112736 Match Control Ulcer Col-M	2.7	113668 Synovium4 Normal	31.6

| 112423 Psoriasis-F | 23.2 | 113669 Syn Fluid Cells4 Normal | 47.0 |

Table HD. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag1355, Run 206231508	Rel. Exp.(%) Ag4501, Run 224702755	Tissue Name	Rel. Exp.(%) Ag1355, Run 206231508	Rel. Exp.(%) Ag4501, Run 224702755
AD 1 Hippo	2.8	2.7	Control (Path) 3 Temporal Ctx	0.0	0.0
AD 2 Hippo	16.6	11.6	Control (Path) 4 Temporal Ctx	7.5	0.0
AD 3 Hippo	6.4	2.5	AD 1 Occipital Ctx	2.5	2.8
AD 4 Hippo	0.0	0.0	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 Hippo	8.7	20.4	AD 3 Occipital Ctx	5.0	1.6
AD 6 Hippo	59.0	41.5	AD 4 Occipital Ctx	2.8	1.9
Control 2 Hippo	3.0	7.0	AD 5 Occipital Ctx	2.6	9.7
Control 4 Hippo	0.0	2.0	AD 6 Occipital Ctx	5.4	2.6
Control (Path) 3 Hippo	4.8	2.9	Control 1 Occipital Ctx	0.0	2.7
AD 1 Temporal Ctx	4.6	5.8	Control 2 Occipital Ctx	3.0	3.3
AD 2 Temporal Ctx	2.2	5.9	Control 3 Occipital Ctx	3.2	0.0
AD 3 Temporal Ctx	7.7	0.0	Control 4 Occipital Ctx	0.0	0.0
AD 4 Temporal	0.0	4.7	Control (Path) 1	8.1	0.9

314

Ctx			Occipital Ctx		
AD 5 Inf Temporal Ctx	4.4	11.7	Control (Path) 2 Occipital Ctx	0.0	1.7
AD 5 Sup Temporal Ctx	7.7	3.5	Control (Path) 3 Occipital Ctx	4.1	0.0
AD 6 Inf Temporal Ctx	74.7	58.6	Control (Path) 4 Occipital Ctx	2.7	0.0
AD 6 Sup Temporal Ctx	**100.0**	**100.0**	Control 1 Parietal Ctx	0.0	2.4
Control 1 Temporal Ctx	0.0	2.8	Control 2 Parietal Ctx	3.5	4.3
Control 2 Temporal Ctx	2.0	3.8	Control 3 Parietal Ctx	2.6	1.0
Control 3 Temporal Ctx	0.0	0.0	Control (Path) 1 Parietal Ctx	11.3	9.2
Control 3 Temporal Ctx	1.9	0.8	Control (Path) 2 Parietal Ctx	6.3	6.5
Control (Path) 1 Temporal Ctx	7.2	1.4	Control (Path) 3 Parietal Ctx	5.8	11.7
Control (Path) 2 Temporal Ctx	0.0	4.3	Control (Path) 4 Parietal Ctx	6.3	1.5

Table HE. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag1355, Run 213323381	Rel. Exp.(%) Ag1355, Run 222654254	Rel. Exp.(%) Ag4501, Run 222695219	Tissue Name	Rel. Exp.(%) Ag1355, Run 213323381	Rel. Exp.(%) Ag1355, Run 222654254	Rel. Exp.(%) Ag4501, Run 222695219
Adipose	43.5	48.6	82.4	Renal ca. TK-10	1.6	2.2	1.9
Melanoma* Hs688(A).T	1.8	2.0	3.5	Bladder	3.4	3.1	3.2
Melanoma* Hs688(B).T	1.2	1.0	2.3	Gastric ca. (liver met.) NCI-N87	0.7	3.2	3.0
Melanoma* M14	0.1	0.6	1.9	Gastric ca. KATO III	0.7	0.4	3.3
Melanoma* LOXIMVI	2.4	0.9	1.0	Colon ca. SW-948	0.0	0.4	1.1
Melanoma* SK-MEL-5	0.3	1.1	2.9	Colon ca. SW480	2.6	3.6	2.4
Squamous cell carcinoma SCC-4	0.0	1.6	3.8	Colon ca.* (SW480 met) SW620	0.3	0.4	0.0
Testis Pool	11.2	20.4	25.5	Colon ca. HT29	0.0	1.0	0.0
Prostate ca.* (bone met) PC-3	2.2	2.1	2.8	Colon ca. HCT-116	12.3	15.0	32.3
Prostate Pool	0.0	0.0	0.5	Colon ca. CaCo-2	1.4	1.0	3.4
Placenta	67.8	**100.0**	98.6	Colon cancer tissue	0.4	2.4	8.5
Uterus Pool	0.5	0.6	3.3	Colon ca. SW1116	0.5	1.0	4.0
Ovarian ca. OVCAR-3	1.5	0.8	1.8	Colon ca. Colo-205	83.5	0.0	0.0
Ovarian ca. SK-OV-3	**100.0**	55.1	**100.0**	Colon ca. SW-48	0.0	0.4	0.0
Ovarian ca. OVCAR-4	0.5	1.0	1.4	Colon Pool	0.8	0.0	0.7
Ovarian ca. OVCAR-5	3.1	2.2	1.7	Small Intestine Pool	1.6	1.4	10.0
Ovarian ca. IGROV-1	4.7	4.0	4.7	Stomach Pool	85.3	5.4	6.3
Ovarian ca.	0.9	0.7	1.3	Bone Marrow	1.3	2.1	4.5

OVCAR-8				Pool			
Ovary	15.2	17.2	26.8	Fetal Heart	0.8	2.4	3.1
Breast ca. MCF-7	10.5	10.3	21.3	Heart Pool	1.1	0.5	0.7
Breast ca. MDA-MB-231	2.5	1.6	5.5	Lymph Node Pool	0.2	1.7	1.6
Breast ca. BT 549	3.9	4.8	9.5	Fetal Skeletal Muscle	7.8	15.1	13.7
Breast ca. T47D	3.3	3.4	5.1	Skeletal Muscle Pool	6.9	7.4	19.5
Breast ca. MDA-N	0.0	0.4	0.0	Spleen Pool	1.3	0.9	0.0
Breast Pool	81.2	0.5	4.7	Thymus Pool	2.9	5.8	6.4
Trachea	3.5	6.8	7.0	CNS cancer (glio/astro) U87-MG	1.2	1.2	2.6
Lung	2.0	4.7	4.8	CNS cancer (glio/astro) U-118-MG	0.4	0.8	2.1
Fetal Lung	2.9	6.8	5.0	CNS cancer (neuro;met) SK-N-AS	0.6	0.4	1.6
Lung ca. NCI-N417	83.5	0.0	0.0	CNS cancer (astro) SF-539	2.6	1.8	7.9
Lung ca. LX-1	1.8	6.8	1.8	CNS cancer (astro) SNB-75	1.8	2.1	2.8
Lung ca. NCI-H146	1.0	0.3	2.8	CNS cancer (glio) SNB-19	3.1	5.0	7.9
Lung ca. SHP-77	0.6	0.3	3.2	CNS cancer (glio) SF-295	2.2	3.7	6.1
Lung ca. A549	0.0	0.7	0.0	Brain (Amygdala) Pool	0.3	0.0	2.7
Lung ca. NCI-H526	0.0	0.0	0.0	Brain (cerebellum)	0.0	0.0	0.0
Lung ca. NCI-H23	2.9	3.1	10.9	Brain (fetal)	14.0	11.5	26.1
Lung ca. NCI-H460	82.9	1.8	2.2	Brain (Hippocampus) Pool	1.8	2.3	6.2
Lung ca. HOP-62	3.3	6.5	13.5	Cerebral Cortex Pool	1.1	0.0	4.6
Lung ca.	0.2	0.0	0.0	Brain	0.5	1.8	0.8

NCI-H522				(Substantia nigra) Pool			
Liver	0.0	0.0	0.0	Brain (Thalamus) Pool	0.0	4.0	6.4
Fetal Liver	0.1	0.0	1.5	Brain (whole)	85.3	0.0	1.5
Liver ca. HepG2	1.5	3.8	3.4	Spinal Cord Pool	1.7	0.7	1.8
Kidney Pool	0.7	0.4	2.7	Adrenal Gland	3.3	4.9	2.7
Fetal Kidney	2.3	7.5	6.7	Pituitary gland Pool	0.5	0.8	1.6
Renal ca. 786-0	1.7	1.4	6.1	Salivary Gland	0.7	1.9	3.1
Renal ca. A498	0.8	0.8	1.9	Thyroid (female)	0.3	0.0	2.7
Renal ca. ACHN	0.0	0.3	0.0	Pancreatic ca. CAPAN2	5.0	11.7	19.3
Renal ca. UO-31	0.0	0.4	1.5	Pancreas Pool	0.6	1.8	1.1

Table HF. Panel 1.2

Tissue Name	Rel. Exp.(%) Ag1355, Run 134848229	Tissue Name	Rel. Exp.(%) Ag1355, Run 134848229
Endothelial cells	0.0	Renal ca. 786-0	0.1
Heart (Fetal)	0.0	Renal ca. A498	0.1
Pancreas	0.1	Renal ca. RXF 393	0.1
Pancreatic ca. CAPAN 2	0.0	Renal ca. ACHN	0.0
Adrenal Gland	1.7	Renal ca. UO-31	0.0
Thyroid	0.0	Renal ca. TK-10	0.0
Salivary gland	0.5	Liver	0.0
Pituitary gland	0.0	Liver (fetal)	0.0
Brain (fetal)	0.1	Liver ca. (hepatoblast) HepG2	0.1
Brain (whole)	0.0	Lung	0.0
Brain (amygdala)	0.0	Lung (fetal)	0.0
Brain (cerebellum)	0.0	Lung ca. (small cell) LX-1	0.1
Brain (hippocampus)	0.1	Lung ca. (small cell) NCI-H69	0.7
Brain (thalamus)	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Cerebral Cortex	0.1	Lung ca. (large cell)NCI-H460	0.2
Spinal cord	0.0	Lung ca. (non-sm. cell) A549	0.0
glio/astro U87-MG	0.1	Lung ca. (non-s.cell) NCI-H23	0.1
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) HOP-62	0.1
astrocytoma SW1783	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (squam.) SW 900	0.1
astrocytoma SF-539	0.1	Lung ca. (squam.) NCI-H596	0.1
astrocytoma SNB-75	0.0	Mammary gland	0.8
glioma SNB-19	0.5	Breast ca.* (pl.ef) MCF-7	0.6
glioma U251	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0

glioma SF-295	0.0	Breast ca.* (pl. ef) T47D	0.0
Heart	0.1	Breast ca. BT-549	0.1
Skeletal Muscle	0.2	Breast ca. MDA-N	0.0
Bone marrow	0.1	Ovary	0.2
Thymus	0.0	Ovarian ca. OVCAR-3	0.1
Spleen	0.1	Ovarian ca. OVCAR-4	0.1
Lymph node	0.4	Ovarian ca. OVCAR-5	0.1
Colorectal Tissue	0.0	Ovarian ca. OVCAR-8	0.2
Stomach	0.2	Ovarian ca. IGROV-1	1.2
Small intestine	0.4	Ovarian ca. (ascites) SK-OV-3	3.6
Colon ca. SW480	0.0	Uterus	0.0
Colon ca.* SW620 (SW480 met)	0.0	Placenta	100.0
Colon ca. HT29	0.0	Prostate	0.1
Colon ca. HCT-116	0.3	Prostate ca.* (bone met) PC-3	0.1
Colon ca. CaCo-2	0.1	Testis	1.7
Colon ca. Tissue (ODO3866)	0.0	Melanoma Hs688(A).T	0.0
Colon ca. HCC-2998	0.0	Melanoma* (met) Hs688(B).T	0.0
Gastric ca.* (liver met) NCI-N87	0.1	Melanoma UACC-62	0.0
Bladder	0.2	Melanoma M14	0.0
Trachea	0.1	Melanoma LOX IMVI	0.0
Kidney	0.1	Melanoma* (met) SK-MEL-5	0.0
Kidney (fetal)	0.2		

320

Table HG. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag1355, Run 173749571	Tissue Name	Rel. Exp.(%) Ag1355, Run 173749571
Normal Colon	7.1	Kidney Margin (OD04348)	4.9
Colon cancer (OD06064)	4.0	Kidney malignant cancer (OD06204B)	5.3
Colon Margin (OD06064)	2.1	Kidney normal adjacent tissue (OD06204E)	5.6
Colon cancer (OD06159)	0.0	Kidney Cancer (OD04450-01)	0.0
Colon Margin (OD06159)	0.0	Kidney Margin (OD04450-03)	0.0
Colon cancer (OD06297-04)	0.0	Kidney Cancer 8120613	0.0
Colon Margin (OD06297-05)	0.0	Kidney Margin 8120614	0.0
CC Gr.2 ascend colon (ODO3921)	4.7	Kidney Cancer 9010320	5.5
CC Margin (ODO3921)	3.8	Kidney Margin 9010321	0.0
Colon cancer metastasis (OD06104)	0.0	Kidney Cancer 8120607	0.0
Lung Margin (OD06104)	0.0	Kidney Margin 8120608	0.0
Colon mets to lung (OD04451-01)	0.0	Normal Uterus	0.0
Lung Margin (OD04451-02)	4.6	Uterine Cancer 064011	0.0
Normal Prostate	0.0	Normal Thyroid	4.7
Prostate Cancer (OD04410)	0.0	Thyroid Cancer 064010	4.5
Prostate Margin (OD04410)	0.0	Thyroid Cancer A302152	0.0
Normal Ovary	0.0	Thyroid Margin A302153	0.0
Ovarian cancer (OD06283-03)	0.0	Normal Breast	53.2
Ovarian Margin (OD06283-07)	10.1	Breast Cancer (OD04566)	0.0
Ovarian Cancer 064008	0.0	Breast Cancer 1024	4.8

Ovarian cancer (OD06145)	52.1	Breast Cancer (OD04590-01)	0.0
Ovarian Margin (OD06145)	**100.0**	Breast Cancer Mets (OD04590-03)	44.8
Ovarian cancer (OD06455-03)	18.2	Breast Cancer Metastasis (OD04655-05)	17.8
Ovarian Margin (OD06455-07)	0.0	Breast Cancer 064006	6.0
Normal Lung	2.4	Breast Cancer 9100266	0.0
Invasive poor diff. lung adeno (ODO4945-01	0.0	Breast Margin 9100265	0.0
Lung Margin (ODO4945-03)	0.0	Breast Cancer A209073	0.0
Lung Malignant Cancer (OD03126)	0.0	Breast Margin A2090734	4.7
Lung Margin (OD03126)	0.0	Breast cancer (OD06083)	4.2
Lung Cancer (OD05014A)	0.0	Breast cancer node metastasis (OD06083)	22.5
Lung Margin (OD05014B)	0.0	Normal Liver	0.0
Lung cancer (OD06081)	0.0	Liver Cancer 1026	0.0
Lung Margin (OD06081)	0.0	Liver Cancer 1025	0.0
Lung Cancer (OD04237-01)	0.0	Liver Cancer 6004-T	17.6
Lung Margin (OD04237-02)	0.0	Liver Tissue 6004-N	39.2
Ocular Melanoma Metastasis	0.0	Liver Cancer 6005-T	0.0
Ocular Melanoma Margin (Liver)	0.0	Liver Tissue 6005-N	4.8
Melanoma Metastasis	0.0	Liver Cancer 064003	0.0
Melanoma Margin (Lung)	0.0	Normal Bladder	0.0
Normal Kidney	0.0	Bladder Cancer 1023	0.0
Kidney Ca, Nuclear grade 2 (OD04338)	0.0	Bladder Cancer A302173	0.0
Kidney Margin (OD04338)	0.0	Normal Stomach	14.3

Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	Gastric Cancer 9060397	0.0
Kidney Margin (OD04339)	0.0	Stomach Margin 9060396	0.0
Kidney Ca, Clear cell type (OD04340)	8.3	Gastric Cancer 9060395	0.0
Kidney Margin (OD04340)	0.0	Stomach Margin 9060394	5.1
Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer 064005	0.0

Table HH. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag1355, Run 170284813	Rel. Exp.(%) Ag4501, Run 197089606	Tissue Name	Rel. Exp.(%) Ag1355, Run 170284813	Rel. Exp.(%) Ag4501, Run 197089606
Secondary Th1 act	0.0	0.0	HUVEC IL-1beta	0.0	8.4
Secondary Th2 act	5.4	0.0	HUVEC IFN gamma	27.0	9.3
Secondary Tr1 act	0.0	0.0	HUVEC TNF alpha + IFN gamma	0.0	10.2
Secondary Th1 rest	0.0	0.0	HUVEC TNF alpha + IL4	18.9	0.0
Secondary Th2 rest	0.0	0.0	HUVEC IL-11	0.0	16.8
Secondary Tr1 rest	10.3	0.0	Lung Microvascular EC none	0.0	23.2
Primary Th1 act	0.0	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0	8.8
Primary Th2 act	0.0	0.0	Microvascular Dermal EC none	0.0	0.0
Primary Tr1 act	25.2	10.2	Microsvasular Dermal EC TNFalpha + IL-1beta	14.3	0.0
Primary Th1 rest	0.0	0.0	Bronchial epithelium TNFalpha + IL1beta	57.8	52.5
Primary Th2 rest	0.0	0.0	Small airway epithelium none	0.0	6.5
Primary Tr1 rest	0.0	0.0	Small airway epithelium TNFalpha + IL-1beta	35.8	79.0
CD45RA CD4 lymphocyte act	23.0	0.0	Coronery artery SMC rest	0.0	18.8
CD45RO CD4 lymphocyte act	0.0	0.0	Coronery artery SMC TNFalpha + IL-1beta	5.8	0.0

CD8 lymphocyte act	18.0	8.2	Astrocytes rest	30.4	0.0
Secondary CD8 lymphocyte rest	0.0	0.0	Astrocytes TNFalpha + IL-1beta	24.5	12.2
Secondary CD8 lymphocyte act	0.0	0.0	KU-812 (Basophil) rest	23.7	10.1
CD4 lymphocyte none	0.0	0.0	KU-812 (Basophil) PMA/ionomycin	13.8	35.1
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	21.2	CCD1106 (Keratinocytes) none	50.7	64.2
LAK cells rest	0.0	23.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	42.9	30.1
LAK cells IL-2	11.9	0.0	Liver cirrhosis	0.0	0.0
LAK cells IL-2+IL-12	0.0	0.0	NCI-H292 none	24.8	0.0
LAK cells IL-2+IFN gamma	17.8	0.0	NCI-H292 IL-4	10.2	0.0
LAK cells IL-2+ IL-18	0.0	0.0	NCI-H292 IL-9	0.0	0.0
LAK cells PMA/ionomycin	0.0	0.0	NCI-H292 IL-13	2.9	0.0
NK Cells IL-2 rest	0.0	0.0	NCI-H292 IFN gamma	13.6	19.5
Two Way MLR 3 day	9.5	0.0	HPAEC none	36.1	0.0
Two Way MLR 5 day	0.0	0.0	HPAEC TNF alpha + IL-1 beta	4.9	14.8
Two Way MLR 7 day	0.0	0.0	Lung fibroblast none	10.7	10.9
PBMC rest	0.0	0.0	Lung fibroblast TNF alpha + IL-1 beta	46.0	**100.0**
PBMC PWM	0.0	0.0	Lung fibroblast IL-4	24.3	6.1
PBMC PHA-L	0.0	0.0	Lung fibroblast IL-9	0.0	0.0
Ramos (B cell) none	0.0	0.0	Lung fibroblast IL-13	0.0	0.0
Ramos (B cell) ionomycin	0.0	0.0	Lung fibroblast IFN gamma	22.8	0.0

325

B lymphocytes PWM	0.0	0.0	Dermal fibroblast CCD1070 rest	0.0	0.0
B lymphocytes CD40L and IL-4	24.7	0.0	Dermal fibroblast CCD1070 TNF alpha	35.8	34.4
EOL-1 dbcAMP	0.0	0.0	Dermal fibroblast CCD1070 IL-1 beta	51.4	11.5
EOL-1 dbcAMP PMA/ionomycin	0.0	0.0	Dermal fibroblast IFN gamma	40.1	89.5
Dendritic cells none	0.0	0.0	Dermal fibroblast IL-4	**100.0**	97.3
Dendritic cells LPS	0.0	0.0	Dermal Fibroblasts rest	52.5	68.8
Dendritic cells anti-CD40	0.0	15.4	Neutrophils TNFa+LPS	11.7	0.0
Monocytes rest	0.0	0.0	Neutrophils rest	24.1	15.8
Monocytes LPS	79.6	40.9	Colon	24.8	27.4
Macrophages rest	0.0	0.0	Lung	0.0	0.0
Macrophages LPS	0.0	32.1	Thymus	62.9	38.7
HUVEC none	0.0	11.5	Kidney	0.0	0.0
HUVEC starved	0.0	18.2			

AI_comprehensive panel_v1.0 Summary: Ag1355 Low to moderate levels of expression of the CG95250-01 gene are detected in most of the samples used in this panel, with highest expression in a psoriasis sample (CT=27). Significant expression of this gene is also detected in bone, cartilage, synovium and synovial fluid samples, normal lung samples, COPD lung, emphysema, atopic asthma, asthma, allergy, Crohn's disease (normal matched control and diseased), ulcerative colitis(normal matched control and diseased), and psoriasis (normal matched control and diseased). Therefore, therapeutic modulation of this gene product may ameliorate symptoms/conditions associated with autoimmune and inflammatory disorders including psoriasis, allergy, asthma, inflammatory bowel disease, rheumatoid arthritis and osteoarthritis

CNS_neurodegeneration_v1.0 Summary: Ag1355/Ag4501 Two experiment with different probe and primer sets are in excellent agreement with highest expression of the CG95250-01 gene in a temporal cortex sample derived from an Alzheimer's disease patient (CT=31.5). This gene is found to be slighltly upregulated in the temporal cortex of

Alzheimer's disease patients. Therefore, therapeutic modulation of this gene product may be of useful in the treatment of this disease and decrease neuronal death.

General_screening_panel_v1.4 Summary: Ag1355/Ag4501 Two experiment with different probe and primer sets are in excellent agreement with highest expression of the CG95250-01 gene in placenta and an ovarian cancer cell line (CTs=30). Therefore, therapeutic modulation of this gene product may be useful in treatment of reproductive disorders and ovarian cancer.

In addition, significant expression of this gene is seen in a ovarian cancer, breast cancer, lung cancer, pancreatic cancer, and colon cancer cell lines. The CG95250-01 gene codes for aminopeptidase N (APN) like protein. Recently, APN has shown to play a role in cell motility and angiogenesis, and it is a useful indicator of a poor prognosis for node-positive patients with colon cancer (Hashida et al., 2002, Gastroenterology 2002 Feb;122(2):376-86, PMID: 11832452). Therefore, therapeutic modulation of the protein encoded by this gene, through the use of small molecule drugs, protein therapeutics or antibodies, could be useful in the treatement of these cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in adipose, adrenal gland, skeletal muscle, and stomach. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

Results from one experiment (run 213323381) with the CG95250-01 gene are not included. The amp plot indicates that there were experimental difficulties with this run.

Panel 1.2 Summary: Ag1355 Highest expression of the CG95250-01 gene is seen in placenta (CT=22). In addition, significant expression of this gene is seen in a ovarian cancer, breast cancer, lung cancer, pancreatic cancer, prostate cancer, renal cancer, CNS cancer, melanoma and colon cancer cell lines. Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in pancrease, liver, heart, adrenal gland, skeletal muscle, small intestine and stomach. Please see panel 1.4 for discussion on the potential utility of this gene.

In addition, this gene is expressed at high levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

327

Panel 2.2 Summary: Ag1355 Highest expression of the CG95250-01 gene is seen in ovarian margin sample (CT=32.7). Low but significant expression of this gene is also seen in ovarian cancer, normal breast and the cancer metastasis, and in normal liver samples. Please see panel 1.4 for the discussion of the utility of the gene.

5   **Panel 4.1D Summary:** Ag1355 Highest expression of the CG95250-01 gene is seen in IL-4 treated dermal fibroblast sample (CT=34). In addition, significant expression of this gene is also detected in thymus, TNFalpha + IL1 beta treated bronchial epithelium, LPS treated monocytes and resting dermal fibroblasts. LPS treated monocytes contribute to the innate and specific immunity by migrating to the site of tissue injury and releasing

10   inflammatory cytokines. Cytokine activated epithelial and dermal fibroblast cells contribute to the inflammation process. The CG95250-01 gene codes for aminopeptidase N (APN) like protein. APN is shown to induce chemotactic migration of leukocytes (Tani et al., 2001, J Med Invest 48:133-41). Thus, APN-induced leukocyte chemotaxis and activation may play an important role in immunologic events of inflammatory and allergic

15   diseases.

Ag4501 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

I. CG95430-01: AdipoQ-like

Expression of gene CG95430-01 was assessed using the primer-probe set Ag4020, described in Table IA. Results of the RTQ-PCR runs are shown in Tables IB, IC, ID and IE.

Table IA. Probe Name Ag4020

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cacattgctggggtctattact-3'	22	458	167
Probe	TET-5'-tcacctaccacatcactgttttctcca-3'-TAMRA	27	480	168
Reverse	5'-ttttgaccaaagacacctgaac-3'	22	512	169

Table IB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag4020, Run 212393803	Tissue Name	Rel. Exp.(%) Ag4020, Run 212393803
AD 1 Hippo	14.6	Control (Path) 3 Temporal Ctx	1.8
AD 2 Hippo	11.8	Control (Path) 4 Temporal Ctx	4.9
AD 3 Hippo	6.0	AD 1 Occipital Ctx	2.4
AD 4 Hippo	7.9	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	5.8	AD 3 Occipital Ctx	1.2
AD 6 Hippo	**100.0**	AD 4 Occipital Ctx	1.9
Control 2 Hippo	15.3	AD 5 Occipital Ctx	2.5
Control 4 Hippo	8.3	AD 6 Occipital Ctx	1.7
Control (Path) 3 Hippo	3.9	Control 1 Occipital Ctx	1.5
AD 1 Temporal Ctx	4.9	Control 2 Occipital Ctx	3.2
AD 2 Temporal Ctx	6.5	Control 3 Occipital Ctx	3.1
AD 3 Temporal Ctx	1.5	Control 4 Occipital Ctx	0.5
AD 4 Temporal Ctx	8.2	Control (Path) 1 Occipital Ctx	14.3
AD 5 Inf Temporal Ctx	12.3	Control (Path) 2 Occipital Ctx	1.0
AD 5 Sup Temporal Ctx	28.3	Control (Path) 3 Occipital Ctx	0.5
AD 6 Inf Temporal Ctx	12.9	Control (Path) 4 Occipital Ctx	3.0
AD 6 Sup Temporal Ctx	8.0	Control 1 Parietal Ctx	0.9
Control 1 Temporal Ctx	0.6	Control 2 Parietal Ctx	6.0
Control 2 Temporal Ctx	2.0	Control 3 Parietal Ctx	3.9
Control 3 Temporal Ctx	3.8	Control (Path) 1 Parietal Ctx	8.0
Control 3	0.0	Control (Path) 2	3.4

Temporal Ctx		Parietal Ctx	
Control (Path) 1 Temporal Ctx	5.3	Control (Path) 3 Parietal Ctx	1.2
Control (Path) 2 Temporal Ctx	2.9	Control (Path) 4 Parietal Ctx	3.6

Table IC. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4020, Run 171614122	Tissue Name	Rel. Exp.(%) Ag4020, Run 171614122
Secondary Th1 act	2.4	HUVEC IL-1beta	1.1
Secondary Th2 act	10.7	HUVEC IFN gamma	0.0
Secondary Tr1 act	1.5	HUVEC TNF alpha + IFN gamma	1.0
Secondary Th1 rest	1.9	HUVEC TNF alpha + IL4	1.0
Secondary Th2 rest	2.5	HUVEC IL-11	0.7
Secondary Tr1 rest	0.0	Lung Microvascular EC none	1.0
Primary Th1 act	1.5	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	3.6	Microvascular Dermal EC none	1.4
Primary Tr1 act	1.6	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	1.3	Bronchial epithelium TNFalpha + IL1beta	5.0
Primary Th2 rest	0.6	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	2.9
CD45RA CD4 lymphocyte act	3.4	Coronery artery SMC rest	1.2
CD45RO CD4 lymphocyte act	3.2	Coronery artery SMC TNFalpha + IL-1beta	2.5
CD8 lymphocyte act	1.4	Astrocytes rest	3.6
Secondary CD8 lymphocyte rest	4.9	Astrocytes TNFalpha + IL-1beta	1.5
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	6.9
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	8.5
LAK cells rest	2.1	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.3
LAK cells IL-2	6.3	Liver cirrhosis	6.6

LAK cells IL-2+IL-12	2.3	NCI-H292 none	2.5
LAK cells IL-2+IFN gamma	2.3	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	2.4	NCI-H292 IL-9	1.8
LAK cells PMA/ionomycin	0.6	NCI-H292 IL-13	5.9
NK Cells IL-2 rest	6.0	NCI-H292 IFN gamma	0.9
Two Way MLR 3 day	1.6	HPAEC none	3.5
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	2.6
Two Way MLR 7 day	0.0	Lung fibroblast none	64.6
PBMC rest	1.1	Lung fibroblast TNF alpha + IL-1 beta	1.8
PBMC PWM	0.0	Lung fibroblast IL-4	25.5
PBMC PHA-L	0.0	Lung fibroblast IL-9	14.8
Ramos (B cell) none	0.0	Lung fibroblast IL-13	26.1
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	33.4
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	3.3
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	4.3
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	2.8
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	1.2
Dendritic cells none	2.9	Dermal fibroblast IL-4	1.4
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	10.5
Dendritic cells anti-CD40	2.6	Neutrophils TNFa+LPS	0.0
Monocytes rest	2.0	Neutrophils rest	1.2
Monocytes LPS	0.0	Colon	3.0
Macrophages rest	2.2	Lung	10.0
Macrophages LPS	2.1	Thymus	19.6
HUVEC none	0.0	Kidney	**100.0**
HUVEC starved	0.9		

Table ID. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag4020, Run 223675497	Tissue Name	Rel. Exp.(%) Ag4020, Run 223675497
97457_Patient-02go_adipose	22.5	94709_Donor 2 AM - A_adipose	2.2
97476_Patient-07sk_skeletal muscle	41.8	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	5.3	94711_Donor 2 AM - C_adipose	0.8
97478_Patient-07pl_placenta	2.7	94712_Donor 2 AD - A_adipose	0.0
99167_Bayer Patient 1	0.0	94713_Donor 2 AD - B_adipose	5.1
97482_Patient-08ut_uterus	5.2	94714_Donor 2 AD - C_adipose	0.0
97483_Patient-08pl_placenta	4.6	94742_Donor 3 U - A_Mesenchymal Stem Cells	1.4
97486_Patient-09sk_skeletal muscle	15.2	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0
97487_Patient-09ut_uterus	21.9	94730_Donor 3 AM - A_adipose	4.3
97488_Patient-09pl_placenta	4.5	94731_Donor 3 AM - B_adipose	3.7
97492_Patient-10ut_uterus	12.2	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	4.5	94733_Donor 3 AD - A_adipose	0.0
97495_Patient-11go_adipose	31.4	94734_Donor 3 AD - B_adipose	1.6
97496_Patient-11sk_skeletal muscle	38.2	94735_Donor 3 AD - C_adipose	0.0
97497_Patient-11ut_uterus	8.4	77138_Liver_HepG2untreated	5.3
97498_Patient-11pl_placenta	2.2	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	45.4	81735_Small Intestine	12.8
97501_Patient-12sk_skeletal muscle	**100.0**	72409_Kidney_Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	15.4	82685_Small intestine_Duodenum	4.4
97503_Patient-12pl_placenta	6.0	90650_Adrenal_Adrenocortical adenoma	0.0

94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	0.0
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.9	72411_Kidney_HRE	0.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	1.3	73139_Uterus_Uterine smooth muscle cells	5.5

Table IE. general oncology screening panel_v_2.4

Tissue Name	Rel. Exp.(%) Ag4020, Run 259744763	Tissue Name	Rel. Exp.(%) Ag4020, Run 259744763
Colon cancer 1	20.0	Bladder cancer NAT 2	1.8
Colon cancer NAT 1	1.1	Bladder cancer NAT 3	1.1
Colon cancer 2	11.0	Bladder cancer NAT 4	27.4
Colon cancer NAT 2	8.0	Adenocarcinoma of the prostate 1	8.9
Colon cancer 3	27.4	Adenocarcinoma of the prostate 2	12.3
Colon cancer NAT 3	49.0	Adenocarcinoma of the prostate 3	20.9
Colon malignant cancer 4	28.1	Adenocarcinoma of the prostate 4	4.6
Colon normal adjacent tissue 4	4.6	Prostate cancer NAT 5	11.6
Lung cancer 1	3.4	Adenocarcinoma of the prostate 6	37.9
Lung NAT 1	3.2	Adenocarcinoma of the prostate 7	24.7
Lung cancer 2	68.8	Adenocarcinoma of the prostate 8	5.5
Lung NAT 2	8.2	Adenocarcinoma of the prostate 9	17.8
Squamous cell carcinoma 3	9.7	Prostate cancer NAT 10	29.7
Lung NAT 3	2.1	Kidney cancer 1	8.7
metastatic melanoma 1	33.4	KidneyNAT 1	6.3
Melanoma 2	5.3	Kidney cancer 2	82.4
Melanoma 3	11.3	Kidney NAT 2	18.4
metastatic melanoma 4	40.3	Kidney cancer 3	7.3
metastatic melanoma 5	100.0	Kidney NAT 3	7.1
Bladder cancer 1	10.3	Kidney cancer 4	8.5
Bladder cancer NAT 1	0.0	Kidney NAT 4	5.5
Bladder cancer 2	5.5		

CNS_neurodegeneration_v1.0 Summary: Ag4020 This panel does not show differential expression of the CG95430-01 gene in Alzheimer's disease. However, this

expression profile confirms the presence of this gene in the brain, with highest expression in the hippocampus of an Alzheimer's patient (CT=31.4). Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia,

5   multiple sclerosis, stroke and epilepsy.

General_screening_panel_v1.4 Summary: Ag4020 Results from one experiment with the CG95430-01 gene are not included. The amp plot indicates that there were experimental difficulties with this run.

Panel 4.1D Summary: Ag4020 The CG95430-01 gene is most highly expressed

10   in kidney (CT=32.3). Low but significant levels of expression are also seen in untreated and cytokine activated lung fibroblasts, and thymus. This expression profile suggests that this gene may be involved in the homeostasis of the lung, thymus, and kidney. Expression of this gene appears to be slightly downregulated in cytokine activated lung fibroblasts suggesting that modulation of this gene product may help to maintain or restore function

15   to the lung during inflammation.

Panel 5 Islet Summary: Ag4020 The CG95430-01 gene is expressed in adipose and skeletal muscle (CTs=31,8-34). This gene encodes a putative adiponectin (also known as adipocyte complement-related protein (ACRP-30), AdipoQ, apM1 (adipose most abundant transcript 1) or GBP28 (28 kDa gelatin binding protein)), a member of the C1q

20   family. This protein is induced over 100-fold in adipocyte differentiation (Scherer et al., J Biol Chem 1995 Nov 10;270(45):26746-9) and is involved in adipocyte signaling (Hu et al., J Biol Chem 1996 May 3;271(18):10697-703). Like other members of the C1q family, it forms a homotrimer and the crystal structure indicates that it likely arose from tumor necrosis factor (TNF; Shapiro and Scherer,Curr Biol 1998 Mar 12;8(6):335-8). Ionomycin

25   increases expression of adiponectin and dibutyryl cAMP and TNF-alpha reduce expression and secretion in 3T3-L1 adipocytes (Kappes and Loffler, Horm Metab Res 2000 Nov-Dec;32(11-12):548-54). Levels of adiponectin are decreased in obese humans (Arita et al., Biochem Biophys Res Commun 1999 Apr 2;257(1):79-83) and mice (Hu et al., J Biol Chem 1996 May 3;271(18):10697-703). A proteolytic cleavage product of

30   adiponectin is reported to increase fatty acid oxidation in muscle and causes weight loss in mice. (Fruebis et al., Proc Natl Acad Sci U S A 2001 Feb 13;98(4):2005-10). A missense mutation in the protein was correlated with a markedly low plasma adiponectin level (Takahashi et al., Int J Obes Relat Metab Disord 2000 Jul;24(7):861-8). Recent papers

have shown that adiponectin reverses insulin resistance in mouse models of lipoatrophy and obesity (Yamauchi et al., Nature Med 2000; 7(8): 941-6), and that it enhances insulin action on the liver (Berg et al., ibid, 947-53). In addition, circulating levels of adiponectin have been shown to be lower in obese than in lean subjects and lower in diabetic patients

5   than in non-diabetic patients, with particularly low levels in subjects with coronary artery disease. Furthermore, in patients who were subjected to a weight loss program that resulted in a 10% reduction of their body mass index, circulating adiponectin levels increased significantly. (Berg AH. Trends Endocrinol Metab. 2002 Mar;13(2):84-9) Therefore based on its homology to adiponectin and its expression profile, this protein

10   may function as a potential therapeutic for the treatment of obesity, type II diabetes and/or their secondary complications.

       Adiponectin also seems to have additional cardiovascular and immune system effects. Levels of this protein are reduced in a cohort of Japanese patients with coronary artery disease (CAD), which correlates with the modulation of endothelial adhesion

15   molecules on treatment of human aortic endothelial cells with adiponectin (Ouchi et al., Circulation 1999 Dec 21-28;100(25):2473-6). This protein is found adhering to vascular walls after injury (Okamoto et al. Horm Metab Res 2000 Feb;32(2):47-50) and presence of adiponectin suppresses the macrophage to foam cell transformation (Ouchi et al., Circulation 2001 Feb 27;103(8):1057-63).In addition, levels of adiponectin were lower in

20   diabetic subjects with CAD relative to non-diabetic subjects or diabetic subjects without CAD (Hotta et al., Arterioscler Thromb Vasc Biol 2000 Jun;20(6):1595-9), indicating that lower levels of adiponectin may be an indicator of macroangiopathy in diabetes. Moreover, this protein negatively regulates the growth of myelomonocytic precursors (in part by inducing apoptosis) and macrophage function (Yokota et al., Blood 2000 Sep

25   1;96(5):1723-32). This effect seems to be via the complement 1Q receptor C1qRp.

       The C1q family of proteins involves members such as the complement subunit C1q, gliacolin, C1q-related protein, cerebellin, CORS26 etc., all of which are secreted. They show the presence of a common domain, the C1q domain, at the C terminus and collagen triple helix repeats at the C terminus. The repeats enable the proteins to form

30   homotrimers and possibly oligomers. Members of this family have been implicated in tissue differentiation, immune regulation, energy homeostasis, synaptic function and in diseases such as obesity and neurodegeneration. Therefore, therapeutic modulation of the expression or function of this gene through the use of monoclonal antibodies may be

useful in the prevention and/or treatment of obesity and diabetes. Furthermore, development of human monoclonal antibodies which inhibit this Adipo-Q like protein may also prove useful in the therapeutic treatment of cachexia that occurs in many forms of cancer.

5   References:

Biol Chem 1995 Nov 10;270(45):26746-26749, A novel serum protein similar to C1q, produced exclusively in adipocytes. Scherer PE, Williams S, Fogliano M, Baldini G, Lodish HF.

We describe a novel 30-kDa secretory protein, Acrp30 (adipocyte complement-
10   related protein of 30 kDa), that is made exclusively in adipocytes and whose mRNA is induced over 100-fold during adipocyte differentiation. Acrp30 is structurally similar to complement factor C1q and to a hibernation-specific protein isolated from the plasma of Siberian chipmunks; it forms large homo-oligomers that undergo a series of post-translational modifications. Like adipsin, secretion of Acrp30 is enhanced by insulin, and
15   Acrp30 is an abundant serum protein. Acrp30 may be a factor that participates in the delicately balanced system of energy homeostasis involving food intake and carbohydrate and lipid catabolism. Our experiments also further corroborate the existence of an insulin-regulated secretory pathway in adipocytes.

J Biol Chem 1996 May 3;271(18):10697-10703, AdipoQ is a novel adipose-
20   specific gene dysregulated in obesity. Hu E, Liang P, Spiegelman BM.

Adipose differentiation is accompanied by changes in cellular morphology, a dramatic accumulation of intracellular lipid and activation of a specific program of gene expression. Using an mRNA differential display technique, we have isolated a novel adipose cDNA, termed adipoQ. The adipoQ cDNA encodes a polypeptide of 247 amino
25   acids with a secretory signal sequence at the amino terminus, a collagenous region (Gly-X-Y repeats), and a globular domain. The globular domain of adipoQ shares significant homology with subunits of complement factor C1q, collagen alpha 1(X), and the brain-specific factor cerebellin. The expression of adipoQ is highly specific to adipose tissue in both mouse and rat. Expression of adipoQ is observed exclusively in mature fat cells as
30   the stromal-vascular fraction of fat tissue does not contain adipoQ mRNA. In cultured 3T3-F442A and 3T3-L1 preadipocytes, hormone-induced differentiation dramatically increases the level of expression for adipoQ. Furthermore, the expression of adipoQ mRNA is significantly reduced in the adipose tissues from obese mice and humans.

Whereas the biological function of this polypeptide is presently unknown, the tissue-specific expression of a putative secreted protein suggests that this factor may function as a novel signaling molecule for adipose tissue.

5 Horm Metab Res 2000 Nov;32(11-12):548-554, Influences of ionomycin, dibutyryl-cycloAMP and tumour necrosis factor-alpha on intracellular amount and secretion of apM1 in differentiating primary human preadipocytes. Kappes A, Loffler G.

3T3-L1-adipocytes produce the adipocyte complement related protein of 30 kD (Acrp30), which is also designated as AdipoQ. In order to study the expression and secretion of the human homologue of this protein, apM1 (adipose Most abundant gene

10 transcript 1, also named gelatin-binding protein of 28 kD [GBP28] or adiponectin), a polyclonal antibody was produced. Both expression and secretion can be detected beginning with day 4 after induction of differentiation. The amount of expressed apM1 correlates with the specific activity of the differentiation marker glycerol-3-phosphate dehydrogenase. Secretion of apM1 is increased by the addition of ionomycin. Both the

15 nonhydrolysable dibutyryl-cycloAMP and tumour necrosis factor alpha reduce the expression and secretion of apM1.

Biochem Biophys Res Commun 1999 Apr 2;257(1):79-83, Paradoxical decrease of an adipose-specific protein, adiponectin, in obesity. Arita Y, Kihara S, Ouchi N, Takahashi M, Maeda K, Miyagawa J, Hotta K, Shimomura I, Nakamura T, Miyaoka K,

20 Kuriyama H, Nishida M, Yamashita S, Okubo K, Matsubara K, Muraguchi M, Ohmoto Y, Funahashi T, Matsuzawa Y.

We isolated the human adipose-specific and most abundant gene transcript, apM1 (Maeda, K., et al., Biochem. Biophys. Res. Commun. 221, 286-289, 1996). The apM1 gene product was a kind of soluble matrix protein, which we named adiponectin. To

25 quantitate the plasma adiponectin concentration, we have produced monoclonal and polyclonal antibodies for human adiponectin and developed an enzyme-linked immunosorbent assay (ELISA) system. Adiponectin was abundantly present in the plasma of healthy volunteers in the range from 1.9 to 17.0 mg/ml. Plasma concentrations of adiponectin in obese subjects were significantly lower than those in non-obese subjects,

30 although adiponectin is secreted only from adipose tissue. The ELISA system developed in this study will be useful for elucidating the physiological and pathophysiological role of adiponectin in humans.

Nat Med 2001 Aug;7(8):941-946, The fat-derived hormone adiponectin reverses insulin resistance associated with both lipoatrophy and obesity. Yamauchi T, Kamon J, Waki H, Terauchi Y, Kubota N, Hara K, Mori Y, Ide T, Murakami K, Tsuboyama-Kasaoka N, Ezaki O, Akanuma Y, Gavrilova O, Vinson C, Reitman ML, Kagechika H,

5   Shudo K, Yoda M, Nakano Y, Tobe K, Nagai R, Kimura S, Tomita M, Froguel P, Kadowaki T.

Adiponectin is an adipocyte-derived hormone. Recent genome-wide scans have mapped a susceptibility locus for type 2 diabetes and metabolic syndrome to chromosome 3q27, where the gene encoding adiponectin is located. Here we show that decreased

10   expression of adiponectin correlates with insulin resistance in mouse models of altered insulin sensitivity. Adiponectin decreases insulin resistance by decreasing triglyceride content in muscle and liver in obese mice. This effect results from increased expression of molecules involved in both fatty-acid combustion and energy dissipation in muscle. Moreover, insulin resistance in lipoatrophic mice was completely reversed by the

15   combination of physiological doses of adiponectin and leptin, but only partially by either adiponectin or leptin alone. We conclude that decreased adiponectin is implicated in the development of insulin resistance in mouse models of both obesity and lipoatrophy. These data also indicate that the replenishment of adiponectin might provide a novel treatment modality for insulin resistance and type 2 diabetes.

20   **general oncology screening panel_v_2.4 Summary:** Ag4020 The CG95430-01 gene is most highly expressed in a metastatic melanoma (CT=32.7). Significant levels of expression are also seen in a lung cancer and a kidney cancer when compared to normal adjacent tissue. Thus, expression of this gene may be useful as a diagnostic marker of the presence of these cancers. Furthermore, therapeutic modulation of the expression or

25   function of this gene may be useful in the treatment of kidney cancer, lung cancer, and melanoma.

J. CG95794-01: TRYPSIN III, CATIONIC PRECURSOR

Expression of gene CG95794-01 was assessed using the primer-probe set Ag4029, described in Table JA. Results of the RTQ-PCR runs are shown in Table JB.

<u>Table JA</u>. Probe Name Ag4029

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ctcctggggctatggttgt-3'	19	734	170
Probe	TET-5'-cctcagaagaacaaacctggagtctaca-3'-TAMRA	28	753	171
Reverse	5'-caatggtctgctgaatccatt-3'	21	802	172

5

Table JB. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4029, Run 171615085	Tissue Name	Rel. Exp.(%) Ag4029, Run 171615085
Secondary Th1 act	1.5	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	1.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.8	Coronery artery SMC rest	2.0
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.7	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	1.0	Liver cirrhosis	0.0

LAK cells IL-2+IL-12	0.0	NCI-H292 none	1.8
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.7	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	2.3
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	1.7
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	1.7
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	8.4
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.9
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	1.4
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	1.1
Monocytes rest	0.0	Neutrophils rest	0.9
Monocytes LPS	0.0	Colon	3.8
Macrophages rest	0.0	Lung	5.7
Macrophages LPS	0.0	Thymus	44.1
HUVEC none	0.0	Kidney	**100.0**
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag4029 Expression of the CG95794-01 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

General_screening_panel_v1.4 Summary: Ag4029 Expression of the CG95794-01 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 4.1D Summary: Ag4029 Significant expression of the CG95794-01 gene, which encodes a trypsin homolog, is limited to the kidney and thymus (CTs=32-33). Administration of trypsin has been shown to decrease the presence of TGF-beta1 in the kidney, a significant factor in the development of diabetic nephropathy (Paczek L. Drugs

5   Exp Clin Res 2001;27(4):141-9). Thus, based on this selective expression profile, expression of this gene could be used as to differentiate between these samples and other samples on this panel and as a marker of these tissues. Furthermore, therapeutic modulation of the expression or function of this gene product may be useful in maintaining or restoring function to these organs during inflammation or disease,

10   specifically diabetes.

K. CG95804-01: KALLIKREIN

Expression of gene CG95804-01 was assessed using the primer-probe set Ag4030, described in Table KA. Results of the RTQ-PCR runs are shown in Table KB.

Table KA. Probe Name Ag4030

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ctcgaattgttggaggatttaa-3'	22	83	173
Probe	TET-5'-agaagaattcccagccctggcaagt-3'-TAMRA	25	110	174
Reverse	5'-gatatttggtgaagcggtacac-3'	22	139	175

15

Table KB. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4030, Run 171615090	Tissue Name	Rel. Exp.(%) Ag4030, Run 171615090
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	0.0

LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	**100.0**
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag4030 Expression of the CG95804-01 gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

General_screening_panel_v1.4 Summary: Ag4030 Expression of the CG95804-01 gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4.1D Summary: Ag4030 Highest expression of the CG95804-01 gene is detected exclusively in IL-13 treated NCI-H292 (CT=33). Thus, expression of this gene can be used to distinguish this sample from other samples used in this panel. The NCI-H292 cell line is a human airway epithelial cell line that produces mucins. Mucus overproduction is an important feature of bronchial asthma and chronic obstructive pulmonary disease samples. The expression of this gene in this mucoepidermoid cell line that is often used as a model for airway epithelium (NCI-H292 cells) suggests that this gene may be important in the proliferation or activation of airway epithelium. Therefore, therapeutics designed with the protein encoded by the transcript may reduce or eliminate symptoms caused by inflammation in lung epithelia in chronic obstructive pulmonary disease, asthma, allergy, and emphysema.

L. CG95861-01: TRANSFORMING GROWTH FACTOR-BETA INDUCED PROTEIN

Expression of gene CG95861-01 was assessed using the primer-probe set Ag2049, described in Table LA. Results of the RTQ-PCR runs are shown in Tables LB, LC, LD, LE, LF and LG.

Table LA. Probe Name Ag2049

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gcatgaccctcacctctatgta-3'	22	610	173
Probe	TET-5'-cagaattccaacatccagatccacca-3'-TAMRA	26	633	174
Reverse	5'-gggcacagttcacagttacaat-3'	22	672	175

Table LB. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag2049, Run 208009950	Tissue Name	Rel. Exp.(%) Ag2049, Run 208009950
Adipose	0.5	Renal ca. TK-10	1.3
Melanoma* Hs688(A).T	72.7	Bladder	0.5
Melanoma* Hs688(B).T	**100.0**	Gastric ca. (liver met.) NCI-N87	0.9
Melanoma* M14	4.2	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.6	Colon ca. SW-948	0.2
Melanoma* SK-MEL-5	1.1	Colon ca. SW480	0.8
Squamous cell carcinoma SCC-4	2.0	Colon ca.* (SW480 met) SW620	1.0
Testis Pool	0.2	Colon ca. HT29	0.1
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.2	Colon ca. CaCo-2	1.0
Placenta	0.8	Colon cancer tissue	2.7
Uterus Pool	0.1	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.2
Ovarian ca. SK-OV-3	3.8	Colon ca. SW-48	1.4
Ovarian ca. OVCAR-4	0.1	Colon Pool	0.6
Ovarian ca. OVCAR-5	6.0	Small Intestine Pool	0.2
Ovarian ca. IGROV-1	0.1	Stomach Pool	0.3
Ovarian ca. OVCAR-8	3.4	Bone Marrow Pool	0.1
Ovary	0.1	Fetal Heart	0.6
Breast ca. MCF-7	0.0	Heart Pool	0.2
Breast ca. MDA-MB-231	1.0	Lymph Node Pool	0.3
Breast ca. BT 549	5.5	Fetal Skeletal Muscle	0.4
Breast ca. T47D	6.1	Skeletal Muscle Pool	0.2
Breast ca. MDA-N	0.5	Spleen Pool	0.2
Breast Pool	0.3	Thymus Pool	0.5
Trachea	0.4	CNS cancer	6.9